B169

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/47, A61K 38/17, C07K 16/18, C12Q 1/68 — A2

(11) International Publication Number: WO 99/41373

(43) International Publication Date: 19 August 1999 (19.08.99)

(21) International Application Number: PCT/US99/02527

(22) International Filing Date: 5 February 1999 (05.02.99)

(30) Priority Data:
09/021,764 — 11 February 1998 (11.02.98) — US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US — 09/021,764 (CIP)
Filed on — 11 February 1998 (11.02.98)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: AU-YOUNG, Janice [US/US]; 1419 Kains Avenue, Berkeley, CA 94702 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive #12, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN TRANSPORT-ASSOCIATED MOLECULES

(57) Abstract

The invention provides human transport-associated molecules (TRANP) and polynucleotides which identify and encode TRANP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of TRANP.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN TRANSPORT-ASSOCIATED MOLECULES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human transport-associated molecules and to the use of these sequences in the diagnosis, treatment, and prevention of cancer and transport disorders.

BACKGROUND OF THE INVENTION

Eukaryotic cells are bound by a lipid bilayer membrane and subdivided into functionally distinct, membrane bound compartments. The membranes maintain the essential differences between the cytosol, the extracellular environment, and the contents of each intercellular organelle. As lipid membranes are highly impermeable to most polar molecules transport of essential nutrients, metabolic waste products, cell signaling molecules, macromolecules and proteins across lipid membranes and between organelles must be mediated by a variety of transport-associated molecules.

Transport of molecules between organelles or vesicles is essential for cell function. Eukaryotic proteins are synthesized within the endoplasmic reticulum (ER), delivered from the ER to the Golgi complex for post-translational processing and sorting, and transported from the Golgi to specific intracellular and extracellular destinations. This intracellular and extracellular transport of protein molecules is termed vesicle trafficking. Trafficking is accomplished by the packaging of protein molecules into specialized vesicles which bud from the donor organelle membrane, are transported, and fuse to the target membrane. Specialized cell types utilize specific vesicle trafficking routes. For instance, in endocrine glands, hormones and other secreted proteins are delivered to secretory granules for exocytosis through the plasma membrane to the cell exterior. In macrophages, peroxidases and proteases are delivered to lysosomes. In fat and muscle cells, glucose transporters are stored in vesicles which fuse with the plasma membrane in response to insulin stimulation.

Numerous proteins are necessary for the formation, targeting, and fusion of transport vesicles and for the proper sorting of proteins into these vesicles. The vesicle

trafficking machinery includes coat proteins which promote the budding of vesicles from donor membranes; vesicle- and target-specific identifiers (v-SNAREs and t-SNAREs) which bind to each other and dock the vesicle to the target membrane; and proteins which bind to SNARE complexes and initiate fusion of the vesicle to the target membrane (SNAPs).

The Rab family of small GTP-binding proteins are key regulators of vesicle transport events. Different Rab proteins are associated with distinct subcellular compartments and their vesicular carriers. GTP binding and hydrolysis of Rab proteins regulates the specificity and direction of vesicle transport. Over 40 Rab proteins have been identified in mammalian cells with sequences that share between 35% and 95% identity indicating a broad range of functional specificities. Usually, these proteins are associated with specific cellular organelles. For example, Rab1 is localized to the ER and Golgi complex, Rab2 in the transitional ER and the cis Golgi network, Rab3 to secretory vesicles, Rab4 to early endosomes, Rab5 to early endosomes and the plasma membrane, Rab6 to medial and trans Golgi cisternae, Rab7 to late endosomes, and Rab9 to late endosomes and the trans Golgi network. In addition, RAB proteins are localized to specific tissue types. RAB17 is found in epithelial cells which contain distinct apical, basolateral, and transcytotic transport pathways, while Rab3a, Rab15, and Rab23 are predominantly expressed in the brain and nervous system. (Kuge, O. et al. (1993) J. Cell. Biol. 123:1727-1734; Olkkonen, V. M. et al. (1994) Gene 138:207-211; Olkkonen, V. M. et al. (1997) Int. Rev. Cytol. 176:1-85; and Stahl, B. (1994) J. Biol. Chem. 269:24770-24776.)

The Golgi apparatus lies at the heart of the vesicle transport pathway, and consists of a stack of cisternae, together with the cis- and trans-Golgi networks on either face of the stack. Molecules are transported through the sequence of compartments that comprise the Golgi apparatus by vesicles which bud from one compartment and fuse with the next. During mitosis, the Golgi apparatus is disassembled into hundreds of vesicles. These fragments can then be partitioned between the two daughter cells. The majority of mitotic Golgi fragments are small uniform vesicles, produced by a coatomer I-dependent mechanism (COPI) which is responsible for both anterograde and retrograde transport through the Golgi stack. There is also a COPI-independent pathway which produces larger, more heterogeneous fragments. It is hypothesized that the accumulation of Golgi

fragments during mitosis is related to inhibition of membrane fusion. The p115 protein has been implicated in three Golgi- trafficking steps: 1) intra-Golgi traffic; 2) endoplasmic reticulum to Golgi traffic; and 3) transcytosis. The p115 has been shown to be involved in a docking step prior to membrane fusion. p115 binding to Golgi membranes is specifically inhibited under mitotic conditions as a result of modifications to the membranes. An early stage of mitotic Golgi disassembly *in vitro* can be inhibited by increasing p115 receptor occupancy with excess p115. These data suggest that the mitotic inhibition of p115 binding contributes to the inhibition of membrane fusion by acting at the early step of vesicle docking. (Sapperstein, S. K. et al. (1996) J. Cell Biol. 132:755-767.)

Vesicles in the process of budding from the ER and the Golgi are covered with a protein coat similar to the clathrin coat of endocytotic vesicles. The protein coat is assembled from cytosolic precursor molecules and is confined to budding regions of the organelle membrane. The coat protein (COP)-coated vesicles are uncoated after budding is complete to allow fusion of the vesicle to the target membrane.

The COP coat consists of two major components, a guanosine triphosphatase and coat protomer (coatomer). The individual proteins are cytosolic until assembled into the protein coat. Coatomer is an equimolar complex of seven proteins, termed alpha-, beta-, beta'-, gamma-, delta-, epsilon- and zeta-COP. The coatomer complex binds to dilysine motifs and reversibly associates with Golgi non-clatherin coated vesicles. These vesicles further mediate transport from the ER , via the Golgi, to the trans-Golgi network. Coatomer complex is required for budding from the Golgi membranes and is essential for retrograde Golgi to ER transport of dilysine motif containing proteins. Polyclonal antibody directed to zeta-COP blocks the binding of coatomer to Golgi membranes and prevents the assembly of COP-coated vesicles on Golgi cisternae. Assay of a conditional lethal, temperature-sensitive epsilon-COP mutant of chinese hamster ovary (CHO) cells indicates that epsilon-COP and associated COP coatomers may be necessary for the establishment or maintenance of Golgi structure, for proper ER-to-Golgi transport of integral membrane and secreted proteins, and for normal endocytotic recycling of low-density lipoprotein receptors. (Kuge, O. et al. (1993) supra; and Cosson, P. et al. (1996) EMBO J. 15:1792-1798).

Transport of proteins and RNA between the nucleus and the cytoplasm occurs

cytoplasmic p53 protein were wild type. In contrast, p53 cDNAs derived from breast cancers with nuclear p53 protein contained a variety of missense mutations and a nonsense mutation. It has been suggested that in some breast cancers, the tumor-suppressing activity of p53 is inactivated by the sequestration of the protein in the cytoplasm, away from its site of action in the cell nucleus. Cytoplasmic wild-type p53 was also found in human cervical carcinoma cell lines. (Moll, U.M. et al. (1992) Proc. Natl. Acad. Sci. USA 89:7262-7266; and Liang, X.H. et al. (1993) Oncogene 8:2645-2652.)

Transport across membranes also depends on transporters, or pumps, membrane-spanning proteins which bind to specific classes of molecules and undergo a series of conformational changes in order to transfer the bound molecule across a membrane. Transport by pumps can occur by a passive, concentration-dependant mechanism or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Examples include facilitative transporters, the secondary active symporters and antiporters driven by ion gradients, and active ATP binding cassette transporters involved in multiple-drug resistance and targeting of antigenic peptides to MHC Class I molecules. Transported substrates range from nutrients and ions to a broad variety of drugs, peptides and proteins.

The energy requirements of most mammalian cells are met through a continuous supply of glucose which circulates in the blood. Glucose enters cells through specific glucose transporter molecules present in the plasma membrane. The family includes passive transporters typical of mammalian tissues and active, H(+)-linked sugar transporters from bacteria. These transporters characteristically contain two groups of six putative membrane-spanning alpha-helices separated by large, hydrophilic, cytoplasmic regions. Both the N-terminal and C-terminal regions of the sequence are also predicted to be cytoplasmic. Biophysical studies on the human erythrocyte glucose transporter indicate that the membrane-spanning alpha-helices associate to form a hydrophilic channel or a substrate-binding cleft extending across the membrane. The mechanism of substrate translocation involves alternate exposure of the substrate-binding site to each face of the membrane via a conformational change. (Pessin, J. E. and Bell G. I. (1992) Annu Rev Physiol 54:911-930.)

Transporters play a major role in the regulation of pH, excretion of drugs, and the cellular K+/Na+ balance. Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the

ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H(+)-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H(+)-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na(+)-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective transporters for organic cations and organic anions in organs including the kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintainence of intercellular pH. (Poole, R. C. and Halestrap, A.P. (1993) Am. J. Physiol. 264:C761-C782; Price, N. T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and Haggstrom, I. (1993) J. Biotechnol. 30: 339-350.)

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport small hydrophilic molecules across biological membranes. They are comprised of two homologous halves, each containing two parts: a transmembrane domain with multiple transmembrane segments and a nucleotide binding domain. Mammalian ABC transporters are found either as complete transporters, e.g. the multiple drug resistance transporter and the cystic fibrosis transmembrane regulator proteins, or as half transporters, e.g., the transporters associated with antigen processing; TAP1 and TAP2 proteins, which dimerize to form the active TAP transporter. Two half ABC transporters have been identified in the human peroxisome membrane: the adrenoleukodystrophy protein (ALDP) and the 70-kDa peroxisomal membrane protein (PMP70). Mutations in the adrenoleukodystrophy gene cause X-linked

adrenoleukodystrophy, an inborn error of peroxisomal β-oxidation of very long chain fatty acids. Mutations in the PMP70 genes have been found in patients with Zellweger syndrome, an inborn error of peroxisome biogenesis. Multidrug resistance (MDR) results from overproduction of another member of the ABC transporter family, P-glycoprotein. MDR is primarily due to increased drug extrusion from the resistant cells by P-glycoprotein. The P-glycoproteins have 2 homologous halves, each with 6 hydrophobic segments adjacent to a consensus sequence for nucleotide binding. The hydrophobic segments may form a membrane channel, whereas the nucleotide binding site may be involved in energization of drug transport. (Saurin, W. et al. (1994) Mol. Microbiol. 12:993-1004; Shani, N., et al. (1996) J. Biol. Chem. 271:8725-8730; and Koster, W., and Bohm, B. (1992) Mol. & Gen. Genet. 232:399-407.)

The etiology of numerous human diseases and disorders can be attributed to defects in the transport of proteins through membranes, to organelles or the cell surface. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g. cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Abnormal hormonal secretion is linked to disorders, e.g., diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotropic hormone). Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease. (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G. M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Eng. J. Med. 332:1475-1480.)

Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include, e.g., fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid syndrome, which includes abdominal cramps, diarrhea, and valvular heart disease, caused by excessive amounts of vasoactive substances, e.g., serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones, secreted from

intestinal tumors. Ectopic synthesis and secretion of biologically active peptides includes, e.g., ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid carcinoma. (Schwartz, M.Z. (1997) Semin. Pediatr. Surg. 3:141-146; and Said, S.I. and Faloona, G.R. (1975) N. Engl. J. Med. 293:155-160.)

The discovery of new human transport-associated molecules and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of cancer and transport disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human transport-associated molecules, referred to collectively as "TRANP" and individually as "TRANP-1", "TRANP-2", "TRANP-3", "TRANP-4", "TRANP-5", "TRANP-6", "TRANP-7", "TRANP-8", and "TRANP-9". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the amino acid sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, or to a fragment of any of these sequences. The invention also provides an isolated and purified polynucleotide sequence encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide sequence encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof.

The invention further provides an isolated and purified polynucleotide which

hybridizes under stringent conditions to the polynucleotide sequence encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof, as well as an isolated and purified polynucleotide sequence which is complementary to the polynucleotide sequence encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide sequence comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, and fragments thereof, as well as an isolated and purified polynucleotide which is complementary to the polynucleotide sequence comprising a polynucleotide selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or fragments thereof, the method comprising the steps of: (a) culturing the host cell containing an

expression vector containing at least a fragment of a polynucleotide sequence encoding TRANP under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide selected from the group consisting of the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, and fragments thereof in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5. SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or fragments thereof, as well as a purified agonist and a purified antagonist to the polypeptide. The invention also provides a method for treating or preventing a cancer associated with increased expression or activity of TRANP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, and fragments thereof.

The invention also provides a method for treating or preventing a transport disorder associated with increased expression or activity of TRANP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, and fragments thereof.

The invention also provides a method for treating or preventing a transport disorder associated with decreased expression or activity of TRANP, the method comprising administering to a subject in need of such treatment an effective amount of the polypeptide selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, and fragments thereof.

The invention also provides a method for detecting a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID

NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, and fragments thereof. in a biological sample containing nucleic acids, the method comprising the steps of (a) hybridizing the complement of the polynucleotide sequence encoding the polypeptide comprising SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, or SEQ ID NO:9, or fragments thereof to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding TRANP in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the

invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRANP," as used herein, refers to the amino acid sequences of substantially purified TRANP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to TRANP, increases or prolongs the duration of the effect of TRANP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of TRANP.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding TRANP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRANP, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same TRANP or a polypeptide with at least one functional characteristic of TRANP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRANP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRANP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRANP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRANP is retained. For example, negatively charged amino acids may include aspartic

acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of TRANP which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of TRANP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art. (See, e.g., Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to TRANP, decreases the amount or the duration of the effect of the biological or immunological activity of TRANP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of TRANP.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fa, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind TRANP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic TRANP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition comprising a given amino acid sequence," as these terms are used herein, refer broadly to

any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding TRANP or fragments of TRANP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW™ Fragment Assembly system (GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding TRANP, by northern analysis is indicative of the presence of nucieic acids encoding TRANP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding TRANP.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of TRANP, of a polynucleotide sequence encoding TRANP, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding TRANP. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains a at least one biological or immunological function of the polypeptide from which it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MegAlign program (DNASTAR, Inc., Madison WI). The MegAlign program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (Higgins, D.G. and P. M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be calculated or counted by the clustal method, or by other methods known in the art, such as the Jotun Hein Method. (See, e.g., Hein, J.

(1990) Methods in Enzymology 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an array of distinct polynucleotides or oligonucleotides arrayed on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of TRANP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of

TRANP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "amplimers," "primers," "oligomers," and "probes," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding TRANP, or fragments thereof, or TRANP itself may comprise a bodily fluid; an extract from a cell, chromosome, organelle,

or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5X SSPE, 0.3% SDS, and 200 μg/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein. describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of TRANP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of new human transport-associated molecules (TRANP), the polynucleotides encoding TRANP, and the use of these compositions for the diagnosis, treatment, or prevention of cancer and transport disorders. Table 1 shows the sequence identification numbers, Incyte Clone identification number, and cDNA library for each of the human transport-associated molecules disclosed herein.

Table 1

PROTEIN	NUCLEOTIDE	CLONE ID	LIBRARY
SEQ ID NO:1	SEQ ID NO:10	144861	TLYMNOR01

SEQ ID NO:2	SEQ ID NO:11	607812	COLNNOT01
SEQ ID NO:3	SEQ ID NO:12	1259384	MENITUT03
SEQ ID NO:4	SEQ ID NO:13	1340813	COLNTUT03
SEQ ID NO:5	SEQ ID NO:14	1689731	PROSTUT10
SEQ ID NO:6	SEQ ID NO:15	2751730	THPIAZS08
SEQ ID NO:7	SEQ ID NO:16	2794975	NPOLNOT01
SEQ ID NO:8	SEQ ID NO:17	2797710	NPOLNOT01
SEQ ID NO:9	SEQ ID NO:18	2914719	THYMFET03

Nucleic acids encoding the TRANP-1 of the present invention were first identified in Incyte Clone 144861 from the mononuclear cell cDNA library (TLYMNOR01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:10, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 144861 (TLYMNOR01) and 156223 (THP1PLB02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1. TRANP-1 is 383 amino acids in length and has two potential sugar transport protein signatures from residues L_{42} to R_{67} and from residues A_{230} to S_{246}. TRANP-1 has chemical and structural homology with a sugar transporter from sugar beet (GI 1209756). In particular, TRANP and GI 1209756; share 30% identity. The fragment of SEQ ID NO:10 from about nucleotide 442 to nucleotide 483 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic/immune, cardiovascular, and urologic cDNA libraries. Approximately 60% of these libraries are associated with fetal tissues and proliferating cell lines and 40% with inflammation.

Nucleic acids encoding the TRANP-2 of the present invention were first identified in Incyte Clone 607812 from the colon tissue cDNA library (COLNNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:11, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 607812 (COLNNOT01), 3176117 (UTRSTUT04), and the shotgun sequences SAEA02352 and SAEC10448.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2. TRANP-2 is 272 amino acids in length and has a potential N-glycosolation site at residue N_{96}, a potential casein kinase II phosphorylation

site at residue T_{48}, four potential protein kinase C phosphorylation sites at residues T_{158}, S_{163}, T_{251}, and S_{265}, and a leucine zipper pattern from residue L_{62} to residue L_{83}. TRANP-2 has chemical and structural homology with the C. elegans ammonium transporter (GI 1125753). In particular, TRANP-2 and GI 1125753 share 47% identity. The fragment of SEQ ID NO:11 from about nucleotide 251 to nucleotide 274 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, gastrointestinal, neural, and urologic cDNA libraries. Approximately 67% of these libraries are associated with cancer, and 28% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-3 of the present invention were first identified in Incyte Clone 1259384 from the meningioma cDNA library (MENITUT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:12, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1259384 (MENITUT03)and 2946062 (BRAITUT23).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3. TRANP-3 is 210 amino acids in length and has a potential amidation site at residue D_{55}, a potential N-glycosolation site at residue N_{83}, five potential casein kinase II phosphorylation sites at residues S_{70}, S_{86}, T_{88}, S_{115}, and S_{199}, four potential protein kinase C phosphorylation sites at residues T_{45}, S_{70}, T_{85}, and S_{199}, and two potential tyrosine phosphorylation sites at residues Y_{44} and Y_{111}. TRANP-3 has chemical and structural homology with the zeta-COP coatomer protein from cow (GI 441486). In particular, TRANP-3 and GI 441486 share 63% identity. The fragment of SEQ ID NO:12 from about nucleotide 63 to nucleotide 125 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, cardiovascular, gastrointestinal, developmental, endocrine, musculoskeletal, and hematopoietic/immune cDNA libraries. Approximately 59% of these libraries are associated with cancer, and 18% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-4 of the present invention were first identified in Incyte Clone 1340813 from the colon tumor cDNA library (COLNTUT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:13, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1340813 and 1343253 (COLNTUT03), 1515390 (PANCTUT01), 758110 (BRAITUT02), 1573508 (LNODNOT03), 1855515 (

HNT3AZT01), and 1971226 (UCMCL5T01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. TRANP-4 is 465 amino acids in length and has chemical and structural homology with the rat monocarboxylate transporter (GI 2463651). In particular, TRANP-4 and GI 2463651 share 90% identity. The fragment of SEQ ID NO:13 from about nucleotide 674 to nucleotide 694 is useful for hybridization. Northern analysis shows the expression of this sequence in hematopoietic/immune, gastrointestinal, reproductive, cardiovascular, neural, musculoskeletal, developmental, urologic, and dermatologic cDNA libraries. Approximately 41% of these libraries are associated with cancer, 32% with inflammation, and 23% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-5 of the present invention were first identified in Incyte Clone 1689731from the prostate tumor cDNA library (PROSTUT10) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:14, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1689731 (PROSTUT10), 1572293 (UTRSNOT05), 1807228 (SINTNOT13), 1556505 (BLADTUT04), 1295734 (PGANNOT03) 1721666 (BLADNOT06), and the shotgun sequences SAEA02030, SAEA00859, and SAEA02913.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. TRANP-5 is 237 amino acids in length and has a potential ATP/GTP-binding site motif from residues G_{16} to S_{23}, and a potential prenylation site from residues C_{234} to $P_{237.}$ TRANP-5 has chemical and structural homology with mouse rab23 (GI 438162). In particular, TRANP-5 and GI 438162 share 93% identity. The fragment of SEQ ID NO:14 from about nucleotide 721 to nucleotide 759 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, cardiovascular, neural, gastrointestinal, urologic, developmental, endocrine, and dermatological cDNA libraries. Approximately 53% of these libraries are associated with cancer, 13% with inflammation, and 13% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-6 of the present invention were first identified in Incyte Clone 2751730 from the mononuclear cell cDNA library (THP1AZSO8) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:15, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2751730 and 2754090 (THP1AZS08), 2174086 (ENDCNOT03),

and 342996 (BRSTNOR01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:6. TRANP-6 is 208 amino acids in length and has a potential ATP/GTP -binding site motif from residues G_{20} to T_{27}. TRANP-6 has chemical and structural homology with human GTP binding protein (GI 550072). In particular, TRANP-6 and GI 550072 share 90% identity. The fragment of SEQ ID NO:15 from about nucleotide 551 to nucleotide 577 is useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, hematopoietic/immune, developmental, and dermatologic cDNA libraries. Approximately 43% of these libraries are associated with fetal tissues and proliferating cell lines, and 28% with cancer.

Nucleic acids encoding the TRANP-7 of the present invention were first identified in Incyte Clone 2794975 from the nasal polyp tissue cDNA library (NPOLNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:16, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2794975 (NPOLNOT01), 3722893 (BRSTNOT23), 2922316 (SININOT04), 2624003 (KERANOT02), 1?94517 (THYRNOT03), 1416892 (BRAINOT12), 1418058 (KIDNNOT09), 008439 (HMC1NOT01), 1873165 (LEUKNOT02), and 1370761 (BSTMNON02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:7. TRANP-7 is 709 amino acids in length and has two potential ABC transporter family signatures from residues F_{328} to L_{339} and from residues L_{611} toT_{625}, and two potential ATP/GTP- binding site motifs from residues G_{210} toT_{217} and from residues G_{525} to S_{532}. TRANP-7 has chemical and structural homology with a C. elegans putative ATP-binding transport protein (GI 500734). In particular, TRANP-7 and GI 500734 share 47% identity. The fragment of SEQ ID NO:16 from about nucleotide 192 to nucleotide 227 is useful for hybridization. Northern analysis shows the expression of this sequence in neural, reproductive, hematopoietic/immune, gastrointestinal, and cardiovascular cDNA libraries. Approximately 38% of these libraries are associated with cancer, 24% with inflammation, and 22% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-8 of the present invention were first identified in Incyte Clone 2797710 from the nasal polyp tissue cDNA library (NPOLNOT01) using a

computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:17, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2797710 (NPOLNOT01), 879491 (THYRNOT02), 234596 (SINTNOT02), 251188 (PANCDIT01), 036294 (HUVENOB01), 2939073 (THYMFET02), 2019482 (CONNNOT01), 2171995 (ENDCNOT03), 517825 (MMLR1DT01), 1714811 (UCMCNOT02), 371259 (LUNGNOT02), 1995866 (BRSTTUT03), 2642178 (LUNGTUT08), and the shotgun sequences SADA01064 and SADA00379.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:8. TRANP-8 is 962 amino acids in length and has five potential N-glycosylation sites at residues N_{101}, N_{123}, N_{243}, N_{451}, and N_{882}, one potential tyrosine kinase phosphorylation site at residue Y_{81}, eighteen potential casein kinase II phosphorylation sites at residues T_{18}, T_{34}, T_{74}, S_{91}, S_{129}, T_{336}, T_{410}, T_{453}, S_{585}, S_{631}, S_{632}, S_{717}, T_{754}, S_{758}, S_{780}, T_{844}, T_{890}, and S_{902}. TRANP-8 has chemical and structural homology with the rat p115 vesicular transport factor (GI 538153). In particular, TRANP-8 and GI 538153 share 93% identity. The fragment of SEQ ID NO:17 from about nucleotide 2688 to nucleotide 2723 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, gastrointestinal, hematopoietic/immune, and cardiovascular cDNA libraries. Approximately 45% of these libraries are associated with cancer, 30% with inflammation, and 17% with fetal tissues and proliferating cell lines.

Nucleic acids encoding the TRANP-9 of the present invention were first identified in Incyte Clone 2914719 from the fetal thymus cDNA library (THYMFET03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:18, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2914719 (THYMFET03), 264268 (HNT2AGT01), 1553167 (BLADTUT04), 550095 (BEPINOT01), 2148915 (BRAINOT09), and 588157 (UTRSNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:9. TRANP-9 is 368 amino acids in length and has two potential beta transducin family Trp-Asp (WD-40) repeats from residues V_{101} to L_{115} and from residues V_{144} to T_{158}. TRANP-9 has chemical and structural homology with C.

elegans polyA RNA export protein (GI 1546734). In particular, TRANP-9 and GI 1546734 share 47% identity. The fragment of SEQ ID NO:18 from about nucleotide 203 to nucleotide 265 is useful for hybridization. Northern analysis shows the expression of this sequence in reproductive, neural, hematopoietic/immune, and gastrointestinal cDNA libraries. Approximately 39% of these libraries are associated with cancer, 33% with inflammation, and 23% with fetal tissues and proliferating cell lines.

The invention also encompasses TRANP variants. A preferred TRANP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the TRANP amino acid sequence, and which contains at least one functional or structural characteristic of TRANP.

The invention also encompasses polynucleotides which encode TRANP. In a particular embodiment, the invention encompasses a polynucleotide consisting of a nucleic acid sequence selected from group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18.

The invention also encompasses a variant of a polynucleotide sequence encoding TRANP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRANP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRANP. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRANP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRANP, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based

on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRANP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRANP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring TRANP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRANP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRANP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRANP and TRANP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRANP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18, and fragments thereof, under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; and Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp., Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading

exonucleases such as those found in the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding TRANP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer complementary to a linker sequence within the vector and a primer specific to the region predicted to encode the gene. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to 72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991)

Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., Genotyper™ and Sequence Navigator™, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRANP may be used in recombinant DNA molecules to direct expression of TRANP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express TRANP.

As will be understood by those of skill in the art, it may be advantageous to produce TRANP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using

methods generally known in the art in order to alter TRANP-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRANP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of TRANP activity, it may be useful to encode a chimeric TRANP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the TRANP encoding sequence and the heterologous protein sequence, so that TRANP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding TRANP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of TRANP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) Proteins, Structures and Molecular Properties, WH Freeman and Co., New York, NY.) Additionally, the amino acid sequence of TRANP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active TRANP, the nucleotide sequences

encoding TRANP or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRANP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, ch. 4, 8, and 16-17; and Ausubel, F.M. et al. (1995, and periodic supplements) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRANP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding TRANP which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, La Jolla, CA) or pSport1™ plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems,

promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding TRANP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for TRANP. For example, when large quantities of TRANP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding TRANP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) Methods Enzymol. 153:516-544.)

In cases where plant expression vectors are used, the expression of sequences encoding TRANP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L.E.

in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.)

An insect system may also be used to express TRANP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding TRANP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding TRANP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which TRANP may be expressed. (See, e.g., Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRANP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing TRANP in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRANP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding TRANP and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation

codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing TRANP can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in *tk*⁻ or *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; and Lowy, I. et al. (1980) Cell 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *npt* confers resistance to the aminoglycosides neomycin and G-418; and *als* or *pat* confer

resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, CA) are also used (See, e.g., Chalfie, M. et al. (1994) Science 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRANP is inserted within a marker gene sequence, transformed cells containing sequences encoding TRANP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRANP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding TRANP and express TRANP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding TRANP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding TRANP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding TRANP to detect transformants containing DNA or RNA encoding TRANP.

A variety of protocols for detecting and measuring the expression of TRANP,

using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRANP is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Section IV; and Maddox, D.E. et al. (1983) J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRANP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRANP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, MI), Promega (Madison, WI), and U.S. Biochemical Corp. (Cleveland, OH). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRANP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRANP may be designed to contain signal sequences which direct secretion of TRANP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding TRANP to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as

histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA), between the purification domain and the TRANP encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing TRANP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281.) The enterokinase cleavage site provides a means for purifying TRANP from the fusion protein. (See, e.g., Kroll, D.J. et al. (1993) DNA Cell Biol. 12:441-453.)

Fragments of TRANP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) Protein: Structures and Molecular Properties, pp. 55-60, W.H. Freeman and Co., New York, NY.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of TRANP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists among the human transport-associated proteins of the invention. In addition, TRANP is expressed in cancer. Therefore, TRANP appears to play a role in cancer and transport disorders. In cancer and transport disorders where expression of TRANP is increased, or where expression of TRANP is promoting cancer and transport disorders, it is desirable to decrease the expression of TRANP. In transport disorders where expression of TRANP is decreased, it is desirable to provide the protein or increase the expression of TRANP.

Therefore, in one embodiment, TRANP or a fragment or derivative thereof may be administered to a subject to treat or prevent a transport disorder associated with decreased expression or activity of TRANP. Such transport disorders include, but are not limited to, adrenoleukodystrophy, cystic fibrosis, cystinuria, cystine nephrolithiasis, Dubin-Johnson

syndrome, glucose-galactose malabsorption syndrome, diabetes mellitus, diabetes insipidus, diastrophic dysplasia, Dubin-Johnson syndrome, Fanconi disease, Fanconi-Bickel syndrome, Hartup disease, hyperbilirubinemia, hypercholesterolemia, hyperinsulinemia, hyper- and hypoglycemia, iminoglycinuria, Grave's disease, goiter, Cushing's disease, and Addison's disease, von Gierk disease, Zellweger syndrome; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia, proliferative glomerulonephritis; inflammatory bowel disease, multidrug resistance, multiple sclerosis; myasthenia gravis, myocardial ischemia, rheumatoid and osteoarthritis, Parkinson's disease, Pendred syndrome, scleroderma, Bartter's and Gitelman's syndromes, Chediak-Higashi and Sjogren's syndromes, Stargardt's disease, systemic lupus erythematosus, and Wilson's disease.

In another embodiment, a vector capable of expressing TRANP or a fragment or derivative thereof may be administered to a subject to treat or prevent a transport disorder including, but not limited to, those listed above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified TRANP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a transport disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRANP may be administered to a subject to treat or prevent a transport disorder associated with increased expression or activity of TRANP.

In one embodiment, an antagonist which modulates the activity of TRANP may be administered to a subject to treat or prevent a cancer associated with increased expression or activity of TRANP. Such cancers can include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds TRANP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a

pharmaceutical agent to cells or tissue which express TRANP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRANP may be administered to a subject to treat or prevent a cancer including, but not limited to, those listed above.

In a further embodiment, an antagonist which modulates the activity of TRANP may be administered to a subject to treat or prevent a transport disorder associated with increased expression or activity of TRANP. Such transport disorders include, but are not limited to, those listed above. In one aspect, an antibody which specifically binds TRANP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express TRANP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRANP may be administered to a subject to treat or prevent a transport disorder including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRANP may be produced using methods which are generally known in the art. In particular, purified TRANP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRANP. Antibodies to TRANP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRANP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species,

various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRANP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of TRANP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRANP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRANP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly

specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; and Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRANP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRANP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRANP epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding TRANP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding TRANP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding TRANP. Thus, complementary molecules or fragments may be used to modulate TRANP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRANP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding TRANP. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding TRANP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding TRANP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding TRANP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRANP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be

evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRANP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of TRANP, antibodies to TRANP, and mimetics, agonists,

antagonists, or inhibitors of TRANP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc,

polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic

solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of TRANP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRANP or fragments thereof, antibodies of TRANP, and agonists, antagonists or inhibitors of TRANP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED50 (the dose therapeutically effective in 50% of the population) or LD50 (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, and it can be expressed as the ED50/LD50 ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide

sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRANP may be used for the diagnosis of disorders characterized by expression of TRANP, or in assays to monitor patients being treated with TRANP or agonists, antagonists, or inhibitors of TRANP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRANP include methods which utilize the antibody and a label to detect TRANP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRANP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRANP expression. Normal or standard values for TRANP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to TRANP under conditions suitable for complex formation The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of TRANP expressed in subject,

control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRANP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of TRANP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRANP, and to monitor regulation of TRANP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRANP or closely related molecules may be used to identify nucleic acid sequences which encode TRANP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding TRANP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the TRANP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:10, or from genomic sequences including promoters, enhancers, and introns of the TRANP gene.

Means for producing specific hybridization probes for DNAs encoding TRANP include the cloning of polynucleotide sequences encoding TRANP or TRANP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase

coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRANP may be used for the diagnosis of a disorder associated with expression of TRANP. Examples of such a disorder include, but are not limited to, transport disorders such as adrenoleukodystrophy, cystic fibrosis, cystinuria, cystine nephrolithiasis, Dubin-Johnson syndrome, glucose-galactose malabsorption syndrome, diabetes mellitus, diabetes insipidus, diastrophic dysplasia, Dubin-Johnson syndrome, Fanconi disease, Fanconi-Bickel syndrome, Hartup disease, hyperbilirubinemia, hypercholesterolemia, hyperinsulinemia, hyper- and hypoglycemia, iminoglycinuria, Grave's disease, goiter, Cushing's disease, and Addison's disease, von Gierk disease, Zellweger syndrome; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia, proliferative glomerulonephritis; inflammatory bowel disease. multidrug resistance, multiple sclerosis; myasthenia gravis, myocardial ischemia, rheumatoid and osteoarthritis, Parkinson's disease, Pendred syndrome, scleroderma, Bartter's and Gitelman's syndromes, Chediak-Higashi and Sjogren's syndromes, Stargardt's disease, systemic lupus erythematosus, and Wilson's disease; and cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRANP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and ELISA assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRANP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRANP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRANP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in

the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRANP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRANP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRANP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRANP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRANP, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRANP, and will be employed under optimized conditions for identification of a specific gene or condition.

Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of TRANP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; and Duplaa, C. et al. (1993) Anal. Biochem. 229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding TRANP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent <u>in situ</u> hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al.

(1995) in Meyers, R.A. (ed.) Molecular Biology and Biotechnology, VCH Publishers New York, NY, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding TRANP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRANP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRANP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with TRANP, or fragments thereof, and washed. Bound TRANP is then detected by methods well known in the art. Purified

TRANP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRANP specifically compete with a test compound for binding TRANP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRANP.

In additional embodiments, the nucleotide sequences which encode TRANP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

For purposes of example, the preparation and sequencing of the PROSTUT10 cDNA library, from which Incyte Clone 1689731 was isolated, is described below. Preparation and sequencing of cDNAs in libraries in the LIFESEQ™ database have varied over time, and the gradual changes involved use of kits, plasmids, and machinery available at the time the library was made and analyzed.

I. PROSTUT10 cDNA Library Construction

The PROSTUT10 cDNA library was constructed from prostate tumor tissue obtained from a 66-year-old Caucasian male. The tissue was excised when the patient underwent a radical prostatectomy and regional lymph node excision. The pathology report indicated a prostate tumor Gleason grade (2+3) involving the left and right side centrally. The tumor was confined and did not involve the capsule. Perineural invasion was absent. Initially, the patient presented with elevated prostate specific antigen (PSA). The patient history included benign hypertension and alcohol use. The patient's family history included a malignant neoplasm of the prostate and a malignant neoplasm of bone and articular cartilage in the patient's father, and benign hypertension the patient's sibling.

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and DNase treated at 37°C. RNA extraction and precipitation were repeated as before. The mRNA was isolated with the Qiagen Oligotex kit (QIAGEN, Inc., Chatsworth, CA) and used to construct the cDNA libraries.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System (Cat. #18248-013, Gibco/BRL, Gaithersburg, MD). PROSTUT10 cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pINCY. The plasmid pINCY was subsequently transformed into DH5a™ competent cells (Cat. #18258-012, Gibco/BRL).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 Plasmid Kit (Catalog #26173, QIAGEN, Inc.). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing

were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S.F. (1993) J. Mol. Evol 36:290-300; and Altschul et al. (1990) J. Mol. Biol. 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) Protein Engineering 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-8} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and other mammalian sequences (mam), and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, CA). This motif analysis program, based on sequence information contained in the Swiss-Prot Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; and Attwood, T. K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.) PROSITE may be used to identify common functional or structural domains in divergent proteins. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a

measure of the chance distribution of the matches.

In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; and Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; and Collin, M. et al. (1993) Protein Sci. 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, F.M. et al. supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding TRANP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V. Extension of TRANP Encoding Polynucleotides

The sequence of one of the polynucleotides of the present invention was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences, Plymouth, MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C for 15 sec

Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37° C, the E. coli mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 μl from each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2 through 4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular

weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

The nucleotide sequences of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, or SEQ ID NO:18 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham, Chicago, IL), and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified using a Sephadex G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, MI). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II (DuPont NEN, Boston, MA).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots to film for several hours, hybridization patterns are compared visually.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize

array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE™. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; and Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the TRANP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRANP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of TRANP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRANP-encoding transcript.

IX. Expression of TRANP

Expression of TRANP is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g., ß-galactosidase upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g.,Sambrook, supra, pp. 404-433; and Rosenberg, M. et al. (1983) Methods Enzymol. 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of TRANP into bacterial growth media which can be used directly in the following assay for activity.

X. Demonstration of TRANP Activity

TRANP transport activity can be demonstrated through the use of a ligand mixing assay that is used to measure transport from early to late endosomal compartments in X. laevis oocytes. Ovaries are dissected from adult female X. laevis anesthetized with 3-aminobenzoic acid ethyl ester (1 g/liter) in ice water, and oocytes are isolated. (Mukhopadhyay A, et al. (1997) J. Cell. Biol. 136(6): 1227-1237). Oocytes are pulsed with 2mg/ml avidin for 5hrs at 18° C, washed, then incubated for 16hrs to allow avidin to transport to a late compartment. The oocytes are then incubated with 1mg/ml biotin-horseradish peroxidase (HRP) for 30 mins at 18° C to label early endocytic compartments. Varying amounts of TRANP are injected into the oocytes, which are incubated at 18° C. Oocytes are collected at several time points after TRANP injection, washed, and lysed in 100µl of phosphate-buffered saline containing 0.3% Triton X-100, 0.2% methylbenzethorium chloride, and 400 µg/ml of BSA-biotin as a scavenger. Finally, the lysates are centrifuged for 30 seconds in a microfuge, and the avidin-biotin complexes are immunoprecipitated using anti-avidin antibody-coated plates by incubation at 4 °C overnight. The plates are then washed at least 5 times to remove unbound proteins. Transport from the early endosomes to the late compartments is quantified by measuring the amount of immunoprecipitated HRP; increased transport due to TRANP is quantitated by comparison with control oocytes.

XI. Production of TRANP Specific Antibodies

TRANP substantially purified using PAGE electrophoresis (see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The TRANP amino acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al. supra, ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XII. Purification of Naturally Occurring TRANP Using Specific Antibodies

Naturally occurring or recombinant TRANP is substantially purified by immunoaffinity chromatography using antibodies specific for TRANP. An immunoaffinity column is constructed by covalently coupling anti-TRANP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRANP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRANP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRANP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRANP is collected.

XIII. Identification of Molecules Which Interact with TRANP

TRANP, or biologically active fragments thereof, are labeled with ^{125}I

Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRANP, washed, and any wells with labeled TRANP complex are assayed. Data obtained using different concentrations of TRANP are used to calculate values for the number, affinity, and association of TRANP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and SEQ ID NO:9, and fragments thereof.

2. A substantially purified variant having at least 90% amino acid identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide sequence of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide sequence of claim 3.

6. An isolated and purified polynucleotide which is complementary to the polynucleotide of claim 3..

7. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, and fragments thereof.

8. An isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide sequence of claim 7.

9. An isolated and purified polynucleotide which is complementary to the polynucleotide of claim 7.

10. An expression vector containing at least a fragment of the polynucleotide of claim 3.

11. A host cell containing the expression vector of claim 10.

12. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, or fragments thereof, the method comprising the steps of:

a) culturing the host cell of claim 11 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

13. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

14. A purified antibody which specifically binds to the polypeptide of claim 1.

15. A purified agonist of the polypeptide of claim 1.

16. A purified antagonist of the polypeptide of claim 1.

17. A method for treating or preventing a transport disorder, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 13.

18. A method for treating or preventing cancer, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 16.

19. A method for treating or preventing a transport disorder, the method comprising administering to a subject in need of such treatment an effective amount of the

antagonist of claim 16.

20. A method for detecting a polynucleotide encoding a polypeptide having the amino acid sequence selected from the group SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18, and fragments thereof, in a biological sample containing nucleic acids, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample.

21. The method of claim 20 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS. INC.
AU-YOUNG, Janice
HILLMAN, Jennifer L.
LAL, Preeti
GUEGLER, Karl J.
CORLEY, Neil C.
YUE, Henry
BANDMAN, Olga
BAUGHN, Mariah R.

<120> HUMAN TRANSPORT-ASSOCIATED MOLECULES

<140> To Be Assigned
<141> Herewith

<150> 09/021,764
<151> 1998-02-11

<160> 18

<170> PERL PROGRAM

<210> 1
<211> 383
<212> PRT
<213> Homo sapiens

<220> -
<223> 144861

<400> 1
Met Ile Leu Asn Asp Leu Leu Gly Arg Lys Leu Ser Ile Met Phe
1 5 10 15
Ser Ala Val Pro Ser Ala Ala Gly Tyr Ala Leu Met Ala Gly Ala
20 25 30
His Gly Leu Trp Met Leu Leu Leu Gly Arg Thr Leu Thr Gly Phe
35 40 45
Ala Gly Gly Leu Thr Ala Ala Cys Ile Pro Val Tyr Val Ser Glu
50 55 60
Ile Ala Pro Pro Gly Val Arg Gly Ala Leu Gly Ala Thr Pro Gln
65 70 75
Leu Met Ala Val Phe Gly Ser Leu Ser Leu Tyr Ala Leu Gly Leu
80 85 90
Leu Leu Pro Trp Arg Trp Leu Ala Val Ala Gly Glu Ala Pro Val
95 100 105
Leu Ile Met Ile Leu Leu Leu Ser Phe Met Pro Asn Ser Pro Arg
110 115 120
Phe Leu Leu Ser Arg Gly Arg Asp Glu Glu Ala Leu Arg Ala Leu
125 130 135
Ala Trp Leu Arg Gly Thr Asp Val Asp Val His Trp Glu Phe Glu
140 145 150
Gln Ile Gln Asp Asn Val Arg Arg Gln Ser Ser Arg Val Ser Trp
155 160 165
Ala Glu Ala Arg Ala Pro His Val Cys Arg Pro Ile Thr Val Ala
170 175 180

```
Leu Leu Met Arg Leu Leu Gln Gln Leu Thr Gly Ile Thr Pro Ile
                185                 190                 195
Leu Val Tyr Leu Gln Ser Ile Phe Asp Ser Thr Ala Val Leu Leu
                200                 205                 210
Pro Pro Lys Asp Asp Ala Ala Ile Val Gly Ala Val Arg Leu Leu
                215                 220                 225
Ser Val Leu Ile Ala Ala Leu Thr Met Asp Leu Ala Gly Arg Lys
                230                 235                 240
Val Leu Leu Phe Val Ser Gly His Leu Ser Val His Ser Gly His
                245                 250                 255
His Val Cys Cys Gln Pro Asp Ser Gly Ala Val His Pro Leu Trp
                260                 265                 270
Pro Gln Ala Ser Glu Pro Gln Gln His Cys Gly Pro Gly Lys Arg
                275                 280                 285
Val Leu Gly Gly Leu Gly Ala Ala Leu Ala Ala Pro Ala Gly Tyr
                290                 295                 300
Leu Thr Leu Val Pro Leu Leu Ala Thr Met Leu Phe Ile Met Gly
                305                 310                 315
Arg Cys Gly Gly Gly Ser Glu Gly Gln Ala Val Leu Gly Val Lys
                320                 325                 330
Gly Trp Val Met Cys Leu Gly Trp Leu Glu Arg Gly Val Ser Ser
                335                 340                 345
Ser Ser Ala Glu Thr Asp Thr Ala Ala Ser Ser His Pro Thr Gly
                350                 355                 360
Pro Glu Pro Ala Ser Ser Ser Glu Gln Ala Thr Pro Trp Ala Gly
                365                 370                 375
Val His His Leu Ala Ala His Val
                380
```

```
<210> 2
<211> 272
<212> PRT
<213> Homo sapiens

<220> -
<223> 607812

<400> 2
Met Ala Ser Ser Asn Thr Val Leu Met Arg Leu Val Ala Ser Ala
  1               5                  10                  15
Tyr Ser Ile Ala Gln Lys Ala Gly Met Ile Val Arg Arg Val Ile
                 20                  25                  30
Ala Glu Gly Asp Leu Gly Ile Val Glu Lys Thr Cys Ala Thr Asp
                 35                  40                  45
Leu Gln Thr Lys Ala Asp Arg Leu Ala Gln Met Ser Ile Cys Ser
                 50                  55                  60
Ser Leu Ala Arg Lys Phe Pro Lys Leu Thr Ile Ile Gly Glu Glu
                 65                  70                  75
Leu Val Val Trp Val Asp Pro Leu Asp Gly Thr Lys Glu Tyr Thr
                 80                  85                  90
Glu Gly Leu Leu Asp Asn Val Thr Val Leu Ile Gly Ile Ala Tyr
                 95                 100                 105
Glu Gly Lys Ala Ile Ala Gly Val Ile Asn Gln Pro Tyr Tyr Asn
                110                 115                 120
Tyr Glu Ala Gly Pro Asp Ala Val Leu Gly Arg Thr Ile Trp Gly
```

```
                125                 130                 135
Val Leu Gly Leu Gly Ala Phe Gly Phe Gln Leu Lys Glu Val Pro
                140                 145                 150
Ala Gly Lys His Ile Ile Thr Thr Thr Arg Ser His Ser Asn Lys
                155                 160                 165
Leu Val Thr Asp Cys Val Ala Ala Met Asn Pro Asp Ala Val Leu
                170                 175                 180
Arg Val Gly Gly Ala Gly Asn Lys Ile Ile Gln Leu Ile Glu Gly
                185                 190                 195
Lys Ala Ser Ala Tyr Val Phe Ala Ser Pro Gly Cys Lys Lys Trp
                200                 205                 210
Asp Thr Cys Ala Pro Glu Val Ile Leu His Ala Val Gly Gly Lys
                215                 220                 225
Leu Thr Asp Ile His Gly Asn Val Leu Gln Tyr His Lys Asp Val
                230                 235                 240
Lys His Met Asn Ser Ala Gly Val Leu Ala Thr Leu Arg Asn Tyr
                245                 250                 255
Asp Tyr Tyr Ala Ser Arg Val Pro Glu Ser Ile Lys Asn Ala Leu
                260                 265                 270
Val Pro
```

<210> 3
<211> 210
<212> PRT
<213> Homo sapiens

<220> -
<223> 1259384

<400> 3

```
Met Gln Arg Pro Glu Ala Trp Pro Arg Pro His Pro Gly Glu Gly
 1              5                   10                  15
Ala Ala Ala Ala Gln Ala Gly Gly Pro Ala Pro Pro Ala Arg Ala
                20                  25                  30
Gly Glu Pro Ser Gly Leu Arg Leu Gln Glu Pro Ser Leu Tyr Thr
                35                  40                  45
Ile Lys Ala Val Phe Ile Leu Asp Asn Asp Gly Arg Arg Leu Leu
                50                  55                  60
Ala Lys Tyr Tyr Asp Asp Thr Phe Pro Ser Met Lys Glu Gln Met
                65                  70                  75
Val Phe Glu Lys Asn Val Phe Asn Lys Thr Ser Arg Thr Glu Ser
                80                  85                  90
Glu Ile Ala Phe Phe Gly Gly Met Thr Ile Val Tyr Lys Asn Ser
                95                  100                 105
Ile Asp Leu Phe Leu Tyr Val Val Gly Ser Ser Tyr Glu Asn Glu
                110                 115                 120
Leu Met Leu Met Ser Val Leu Thr Cys Leu Phe Glu Ser Leu Asn
                125                 130                 135
His Met Leu Arg Lys Asn Val Glu Lys Arg Trp Leu Leu Glu Asn
                140                 145                 150
Met Asp Gly Ala Phe Leu Val Leu Asp Glu Ile Val Asp Gly Gly
                155                 160                 165
Val Ile Leu Glu Ser Asp Pro Gln Gln Val Ile Gln Lys Val Asn
                170                 175                 180
Phe Arg Ala Asp Asp Gly Gly Leu Thr Glu Gln Ser Val Ala Gln
```

```
                185                 190                 195
Val Leu Gln Ser Ala Lys Glu Gln Ile Lys Trp Ser Leu Leu Lys
                200                 205                 210
```

<210> 4
<211> 465
<212> PRT
<213> Homo sapiens

<220> -
<223> 1340813

<400> 4

```
Met Gly Gly Ala Val Val Asp Glu Gly Pro Thr Gly Val Lys Ala
 1               5                  10                  15
Pro Asp Gly Gly Trp Gly Trp Ala Val Leu Phe Gly Cys Phe Val
                20                  25                  30
Ile Thr Gly Phe Ser Tyr Ala Phe Pro Lys Ala Val Ser Val Phe
                35                  40                  45
Phe Lys Glu Leu Ile Gln Glu Phe Gly Ile Gly Tyr Ser Asp Thr
                50                  55                  60
Ala Trp Ile Ser Ser Ile Leu Leu Ala Met Leu Tyr Gly Thr Gly
                65                  70                  75
Pro Leu Cys Ser Val Cys Val Asn Arg Phe Gly Cys Arg Pro Val
                80                  85                  90
Met Leu Val Gly Gly Leu Phe Ala Ser Leu Gly Met Val Ala Ala
                95                  100                 105
Ser Phe Cys Arg Ser Ile Ile Gln Val Tyr Leu Thr Thr Gly Val
                110                 115                 120
Ile Thr Gly Leu Gly Leu Ala Leu Asn Phe Gln Pro Ser Leu Ile
                125                 130                 135
Met Leu Asn Arg Tyr Phe Ser Lys Arg Arg Pro Met Ala Asn Gly
                140                 145                 150
Leu Ala Ala Ala Gly Ser Pro Val Phe Leu Cys Ala Leu Ser Pro
                155                 160                 165
Leu Gly Gln Leu Leu Gln Asp Arg Tyr Gly Trp Arg Gly Gly Phe
                170                 175                 180
Leu Ile Leu Gly Gly Leu Leu Leu Asn Cys Cys Val Cys Ala Ala
                185                 190                 195
Leu Met Arg Pro Leu Val Val Thr Ala Gln Pro Gly Ser Gly Pro
                200                 205                 210
Pro Arg Pro Ser Arg Arg Leu Leu Asp Leu Ser Val Phe Arg Asp
                215                 220                 225
Arg Gly Phe Val Leu Tyr Ala Val Ala Ala Ser Val Met Val Leu
                230                 235                 240
Gly Leu Phe Val Pro Pro Val Phe Val Val Ser Tyr Ala Lys Asp
                245                 250                 255
Leu Gly Val Pro Asp Thr Lys Ala Ala Phe Leu Leu Thr Ile Leu
                260                 265                 270
Gly Phe Ile Asp Ile Phe Ala Arg Pro Ala Ala Gly Phe Val Ala
                275                 280                 285
Gly Leu Gly Lys Val Arg Pro Tyr Ser Val Tyr Leu Phe Ser Phe
                290                 295                 300
Ser Met Phe Phe Asn Gly Leu Ala Asp Leu Ala Gly Ser Thr Ala
                305                 310                 315
Gly Asp Tyr Gly Gly Leu Val Val Phe Cys Ile Phe Phe Gly Ile
                320                 325                 330
```

```
Ser Tyr Gly Met Val Gly Ala Leu Gln Phe Glu Val Leu Met Ala
                335                 340                 345
Ile Val Gly Thr His Lys Phe Ser Ser Ala Ile Gly Leu Val Leu
                350                 355                 360
Leu Met Glu Ala Val Ala Val Leu Val Gly Pro Pro Ser Gly Gly
                365                 370                 375
Lys Leu Leu Asp Ala Thr His Val Tyr Met Tyr Val Phe Ile Leu
                380                 385                 390
Ala Gly Ala Glu Val Leu Thr Ser Ser Leu Ile Leu Leu Leu Gly
                395                 400                 405
Asn Phe Phe Cys Ile Arg Lys Lys Pro Lys Glu Pro Gln Pro Glu
                410                 415                 420
Val Ala Ala Val Glu Glu Glu Lys Leu His Lys Pro Pro Ala Asp
                425                 430                 435
Ser Gly Val Asp Leu Arg Glu Val Glu His Phe Leu Lys Ala Glu
                440                 445                 450
Pro Glu Lys Asn Gly Glu Val Val His Thr Pro Glu Thr Ser Val
                455                 460                 465
```

<210> 5
<211> 237
<212> PRT
<213> Homo sapiens

<220> -
<223> 1689731

<400> 5

```
Met Leu Glu Glu Asp Met Glu Val Ala Ile Lys Met Val Val Val
  1               5                  10                  15
Gly Asn Gly Ala Val Gly Lys Ser Ser Met Ile Gln Arg Tyr Cys
                 20                  25                  30
Lys Gly Ile Phe Thr Lys Asp Tyr Lys Lys Thr Ile Gly Val Asp
                 35                  40                  45
Phe Leu Glu Arg Gln Ile Gln Val Asn Asp Glu Asp Val Arg Leu
                 50                  55                  60
Met Leu Trp Asp Thr Ala Gly Gln Glu Glu Phe Asp Ala Ile Thr
                 65                  70                  75
Lys Ala Tyr Tyr Arg Gly Ala Gln Ala Cys Val Leu Val Phe Ser
                 80                  85                  90
Thr Thr Asp Arg Glu Ser Phe Glu Ala Val Ser Ser Trp Arg Glu
                 95                 100                 105
Lys Val Val Ala Glu Val Gly Asp Ile Pro Thr Val Leu Val Gln
                110                 115                 120
Asn Lys Ile Asp Leu Leu Asp Asp Ser Cys Ile Lys Asn Glu Glu
                125                 130                 135
Ala Glu Ala Leu Ala Lys Arg Leu Lys Leu Arg Phe Tyr Arg Thr
                140                 145                 150
Ser Val Lys Glu Asp Leu Asn Val Asn Glu Val Phe Lys Tyr Leu
                155                 160                 165
Ala Glu Lys Tyr Leu Gln Lys Leu Lys Gln Gln Ile Ala Glu Asp
                170                 175                 180
Pro Glu Leu Thr His Ser Ser Ser Asn Lys Ile Gly Val Phe Asn
                185                 190                 195
Thr Ser Gly Gly Ser His Ser Gly Gln Asn Ser Gly Thr Leu Asn
```

200 205 210
Gly Gly Asp Val Ile Asn Leu Arg Pro Asn Lys Gln Arg Thr Lys
215 220 225
Lys Asn Arg Asn Pro Phe Ser Ser Cys Ser Ile Pro
230 235

<210> 6
<211> 208
<212> PRT
<213> Homo sapiens

<220> -
<223> 2751730

<400> 6
Met Ser Ala Arg Gly Asp Phe Gly Asn Pro Leu Arg Lys Phe Lys
1 5 10 15
Leu Val Phe Leu Gly Glu Gln Ser Val Gly Lys Thr Ser Leu Ile
20 25 30
Thr Arg Phe Met Tyr Asp Ser Phe Asp Asn Thr Tyr Gln Ala Thr
35 40 45
Ile Gly Ile Asp Phe Leu Ser Lys Thr Met Tyr Leu Glu Asp Arg
50 55 60
Thr Val Arg Leu Gln Leu Trp Asp Thr Ala Gly Gln Glu Arg Phe
65 70 75
Arg Ser Leu Ile Pro Ser Tyr Ile Arg Asp Ser Thr Val Ala Val
80 85 90
Val Val Tyr Asp Ile Thr Asn Leu Asn Ser Phe Gln Gln Thr Ser
95 100 105
Lys Trp Ile Asp Asp Val Arg Thr Glu Arg Gly Ser Asp Val Ile
110 115 120
Ile Met Leu Val Gly Asn Lys Thr Asp Leu Ala Asp Lys Arg Gln
125 130 135
Ile Thr Ile Glu Glu Gly Glu Gln Arg Ala Lys Glu Leu Ser Val
140 145 150
Met Phe Ile Glu Thr Ser Ala Lys Thr Gly Tyr Asn Val Lys Gln
155 160 165
Leu Phe Arg Arg Val Ala Ser Ala Leu Pro Gly Met Glu Asn Val
170 175 180
Gln Glu Lys Ser Lys Glu Gly Met Ile Asp Ile Lys Leu Asp Lys
185 190 195
Pro Gln Glu Pro Pro Ala Ser Glu Gly Gly Cys Ser Cys
200 205

<210> 7
<211> 709
<212> PRT
<213> Homo sapiens

<220> -
<223> 2794975

<400> 7

Met Ala Thr Cys Ala Glu Ile Leu Arg Ser Glu Phe Pro Glu Ile
1 5 10 15
Asp Gly Gln Val Phe Asp Tyr Val Thr Gly Val Leu His Ser Gly
20 25 30
Ser Ala Asp Phe Glu Ser Val Asp Asp Leu Val Glu Ala Val Gly
35 40 45
Glu Leu Leu Gln Glu Val Ser Gly Asp Ser Lys Asp Asp Ala Gly
50 55 60
Ile Arg Ala Val Cys Gln Arg Met Tyr Asn Thr Leu Arg Leu Ala
65 70 75
Glu Pro Gln Ser Gln Gly Asn Ser Gln Val Leu Leu Asp Ala Pro
80 85 90
Ile Gln Leu Ser Lys Ile Thr Glu Asn Tyr Asp Cys Gly Thr Lys
95 100 105
Leu Pro Gly Leu Leu Lys Arg Glu Gln Ser Ser Thr Val Asn Ala
110 115 120
Lys Lys Leu Glu Lys Ala Glu Ala Arg Leu Lys Ala Lys Gln Glu
125 130 135
Lys Arg Ser Glu Lys Asp Thr Leu Lys Thr Ser Asn Pro Leu Val
140 145 150
Leu Glu Glu Ala Ser Ala Ser Gln Ala Gly Ser Arg Lys Glu Ser
155 160 165
Arg Leu Glu Ser Ser Gly Lys Asn Lys Ser Tyr Asp Val Arg Ile
170 175 180
Glu Asn Phe Asp Val Ser Phe Gly Asp Arg Val Leu Leu Ala Gly
185 190 195
Ala Asp Val Asn Leu Ala Trp Gly Arg Arg Tyr Gly Leu Val Gly
200 205 210
Arg Asn Gly Leu Gly Lys Thr Thr Leu Leu Lys Met Leu Ala Thr
215 220 225
Arg Ser Leu Arg Val Pro Ala His Ile Ser Leu Leu His Val Glu
230 235 240
Gln Glu Val Ala Gly Asp Asp Thr Pro Ala Leu Gln Ser Val Leu
245 250 255
Glu Ser Asp Ser Val Arg Glu Asp Leu Leu Arg Arg Glu Arg Glu
260 265 270
Leu Thr Ala Gln Ile Ala Ala Gly Arg Ala Glu Gly Ser Glu Ala
275 280 285
Ala Glu Leu Ala Glu Ile Tyr Ala Lys Leu Glu Glu Ile Glu Ala
290 295 300
Asp Lys Ala Pro Ala Arg Ala Ser Val Ile Leu Ala Gly Leu Gly
305 310 315
Phe Thr Pro Lys Met Gln Gln Gln Pro Thr Arg Glu Phe Ser Gly
320 325 330
Gly Trp Arg Met Arg Leu Ala Leu Ala Arg Ala Leu Phe Ala Arg
335 340 345
Pro Asp Leu Leu Leu Leu Asp Glu Pro Thr Asn Met Leu Asp Val
350 355 360
Arg Ala Ile Leu Trp Leu Glu Asn Tyr Leu Gln Thr Trp Pro Ser
365 370 375
Thr Ile Leu Val Val Ser His Asp Arg Asn Phe Leu Asn Ala Ile
380 385 390
Ala Thr Asp Ile Ile His Leu His Ser Gln Arg Leu Asp Gly Tyr
395 400 405
Arg Gly Asp Phe Glu Thr Phe Ile Lys Ser Lys Gln Glu Arg Leu
410 415 420
Leu Asn Gln Gln Arg Glu Tyr Glu Ala Gln Gln Gln Tyr Arg Gln

425 430 435
His Ile Gln Val Phe Ile Asp Arg Phe Arg Tyr Asn Ala Asn Arg
440 445 450
Ala Ser Gln Val Gln Ser Lys Leu Lys Met Leu Glu Lys Leu Pro
455 460 465
Glu Leu Lys Pro Val Asp Lys Glu Ser Glu Val Val Met Lys Phe
470 475 480
Pro Asp Gly Phe Glu Lys Phe Ser Pro Pro Ile Leu Gln Leu Asp
485 490 495
Glu Val Asp Phe Tyr Tyr Asp Pro Lys His Val Ile Phe Ser Arg
500 505 510
Leu Ser Val Ser Ala Asp Leu Glu Ser Arg Ile Cys Val Val Gly
515 520 525
Glu Asn Gly Ala Gly Lys Ser Thr Met Leu Lys Leu Leu Leu Gly
530 535 540
Asp Leu Ala Pro Val Arg Gly Ile Arg His Ala His Arg Asn Leu
545 550 555
Lys Ile Gly Tyr Phe Ser Gln His His Val Glu Gln Leu Asp Leu
560 565 570
Asn Val Ser Ala Val Glu Leu Leu Ala Arg Lys Phe Pro Gly Arg
575 580 585
Pro Glu Glu Glu Tyr Arg His Gln Leu Gly Arg Tyr Gly Ile Ser
590 595 600
Gly Glu Leu Ala Met Arg Pro Leu Ala Ser Leu Ser Gly Gly Gln
605 610 615
Lys Ser Arg Val Ala Phe Ala Gln Met Thr Met Pro Cys Pro Asn
620 625 630
Phe Tyr Ile Leu Asp Glu Pro Thr Asn His Leu Asp Met Glu Thr
635 640 645
Ile Glu Ala Leu Gly Arg Ala Leu Asn Asn Phe Arg Gly Gly Val
650 655 660
Ile Leu Val Ser His Asp Glu Arg Phe Ile Arg Leu Val Cys Arg
665 670 675
Glu Leu Trp Val Cys Glu Gly Gly Gly Val Thr Arg Val Glu Gly
680 685 690
Gly Phe Asp Gln Tyr Arg Ala Leu Leu Gln Glu Gln Phe Arg Arg
695 700 705
Glu Gly Phe Leu

<210> 8
<211> 962
<212> PRT
<213> Homo sapiens

<220> -
<223> 2797710

<400> 8
Met Asn Phe Leu Arg Gly Val Met Gly Gly Gln Ser Ala Gly Pro
1 5 10 15
Gln His Thr Glu Ala Glu Thr Ile Gln Lys Leu Cys Asp Arg Val
20 25 30
Ala Ser Ser Thr Leu Leu Asp Asp Arg Arg Asn Ala Val Arg Ala
35 40 45
Leu Lys Ser Leu Ser Lys Lys Tyr Arg Leu Glu Val Gly Ile Gln

```
                50                  55                  60
Ala Met Glu His Leu Ile His Val Leu Gln Thr Asp Arg Ser Asp
                65                  70                  75
Ser Glu Ile Ile Gly Tyr Ala Leu Asp Thr Leu Tyr Asn Ile Ile
                80                  85                  90
Ser Asn Glu Glu Glu Glu Glu Val Glu Glu Asn Ser Thr Arg Gln
                95                 100                 105
Ser Glu Asp Leu Gly Ser Gln Phe Thr Glu Ile Phe Ile Lys Gln
               110                 115                 120
Gln Glu Asn Val Thr Leu Leu Leu Ser Leu Leu Glu Glu Phe Asp
               125                 130                 135
Phe His Val Arg Trp Pro Gly Val Lys Leu Leu Thr Ser Leu Leu
               140                 145                 150
Lys Gln Leu Gly Pro Gln Val Gln Gln Ile Ile Leu Val Ser Pro
               155                 160                 165
Met Gly Val Ser Arg Leu Met Asp Leu Leu Ala Asp Ser Arg Glu
               170                 175                 180
Val Ile Arg Asn Asp Gly Val Leu Leu Leu Gln Ala Leu Thr Arg
               185                 190                 195
Ser Asn Gly Ala Ile Gln Lys Ile Val Ala Phe Glu Asn Ala Phe
               200                 205                 210
Glu Arg Leu Leu Asp Ile Ile Ser Glu Glu Gly Asn Ser Asp Gly
               215                 220                 225
Gly Ile Val Val Glu Asp Cys Leu Ile Leu Leu Gln Asn Leu Leu
               230                 235                 240
Lys Asn Asn Asn Ser Asn Arg Asn Phe Phe Lys Glu Gly Ser Tyr
               245                 250                 255
Ile Gln Arg Met Lys Pro Trp Phe Glu Val Gly Asp Glu Asn Ser
               260                 265                 270
Gly Trp Ser Ala Gln Lys Val Thr Asn Leu His Leu Met Leu Gln
               275                 280                 285
Leu Val Arg Val Leu Val Ser Pro Thr Asn Pro Pro Gly Ala Thr
               290                 295                 300
Ser Ser Cys Gln Lys Ala Met Phe Gln Cys Gly Leu Leu Gln Gln
               305                 310                 315
Leu Cys Thr Ile Leu Met Ala Thr Gly Val Pro Ala Asp Ile Leu
               320                 325                 330
Thr Glu Thr Ile Asn Thr Val Ser Glu Val Ile Arg Gly Cys Gln
               335                 340                 345
Val Asn Gln Asp Tyr Phe Ala Ser Val Asn Ala Pro Ser Asn Pro
               350                 355                 360
Pro Arg Pro Ala Ile Val Val Leu Leu Met Ser Met Val Asn Glu
               365                 370                 375
Arg Gln Pro Phe Val Leu Arg Cys Ala Val Leu Tyr Cys Phe Gln
               380                 385                 390
Cys Phe Leu Tyr Lys Asn Gln Lys Gly Gln Gly Glu Ile Val Ser
               395                 400                 405
Thr Leu Leu Pro Ser Thr Ile Asp Ala Thr Gly Asn Ser Val Ser
               410                 415                 420
Ala Gly Gln Leu Leu Cys Gly Gly Leu Phe Ser Thr Asp Ser Leu
               425                 430                 435
Ser Asn Trp Cys Ala Ala Val Ala Leu Ala His Ala Leu Gln Glu
               440                 445                 450
Asn Ala Thr Gln Lys Glu Gln Leu Leu Arg Val Gln Leu Ala Thr
               455                 460                 465
Ser Ile Gly Asn Pro Pro Val Ser Leu Leu Gln Gln Cys Thr Asn
               470                 475                 480
Ile Leu Ser Gln Gly Ser Lys Ile Gln Thr Arg Val Gly Leu Leu
```

```
                    485                 490                 495
Met Leu Leu Cys Thr Trp Leu Ser Asn Cys Pro Ile Ala Val Thr
                500                 505                 510
His Phe Leu His Asn Ser Ala Asn Val Pro Phe Leu Thr Gly Gln
                515                 520                 525
Ile Ala Glu Asn Leu Gly Glu Glu Glu Gln Leu Val Gln Gly Leu
                530                 535                 540
Cys Ala Leu Leu Leu Gly Ile Ser Ile Tyr Phe Asn Asp Asn Ser
                545                 550                 555
Leu Glu Ser Tyr Met Lys Glu Lys Leu Lys Gln Leu Ile Glu Lys
                560                 565                 570
Arg Ile Gly Lys Glu Asn Phe Ile Glu Lys Leu Gly Phe Ile Ser
                575                 580                 585
Lys His Glu Leu Tyr Ser Arg Ala Ser Gln Lys Pro Gln Pro Asn
                590                 595                 600
Phe Pro Ser Pro Glu Tyr Met Ile Phe Asp His Glu Phe Thr Lys
                605                 610                 615
Leu Val Lys Glu Leu Glu Gly Val Ile Thr Lys Ala Ile Tyr Lys
                620                 625                 630
Ser Ser Glu Glu Asp Lys Lys Glu Glu Glu Val Lys Lys Thr Leu
                635                 640                 645
Glu Gln His Asp Asn Ile Val Thr His Tyr Lys Asn Met Ile Arg
                650                 655                 660
Glu Gln Asp Leu Gln Leu Glu Glu Leu Arg Gln Gln Val Ser Thr
                665                 670                 675
Leu Lys Cys Gln Asn Glu Gln Leu Gln Thr Ala Val Thr Gln Gln
                680                 685                 690
Val Ser Gln Ile Gln Gln His Lys Asp Gln Tyr Asn Leu Leu Lys
                695                 700                 705
Ile Gln Leu Gly Lys Asp Asn Gln His Gln Gly Ser Tyr Ser Glu
                710                 715                 720
Gly Ala Gln Met Asn Gly Ile Gln Pro Glu Glu Ile Gly Arg Leu
                725                 730                 735
Arg Glu Glu Ile Glu Glu Leu Lys Arg Asn Gln Glu Leu Leu Gln
                740                 745                 750
Ser Gln Leu Thr Glu Lys Asp Ser Met Ile Glu Asn Met Lys Ser
                755                 760                 765
Ser Gln Thr Ser Gly Thr Asn Glu Gln Ser Ser Ala Ile Val Ser
                770                 775                 780
Ala Arg Asp Ser Glu Gln Val Ala Glu Leu Lys Gln Glu Leu Ala
                785                 790                 795
Thr Leu Lys Ser Gln Leu Asn Ser Gln Ser Val Glu Ile Thr Lys
                800                 805                 810
Leu Gln Thr Glu Lys Gln Glu Leu Leu Gln Lys Thr Glu Ala Phe
                815                 820                 825
Ala Lys Ser Val Glu Val Gln Gly Glu Thr Glu Thr Ile Ile Ala
                830                 835                 840
Thr Lys Thr Thr Asp Val Glu Gly Arg Leu Ser Ala Leu Leu Gln
                845                 850                 855
Glu Thr Lys Glu Leu Lys Asn Glu Ile Lys Ala Leu Ser Glu Glu
                860                 865                 870
Arg Thr Ala Ile Lys Glu Gln Leu Asp Ser Ser Asn Ser Thr Ile
                875                 880                 885
Ala Ile Leu Gln Thr Glu Lys Asp Lys Leu Glu Leu Glu Ile Thr
                890                 895                 900
Asp Ser Lys Lys Glu Gln Asp Asp Leu Leu Val Leu Leu Ala Asp
                905                 910                 915
Gln Asp Gln Lys Ile Leu Ser Leu Lys Asn Lys Leu Lys Asp Leu
```

```
                920                 925                 930
Gly His Pro Val Glu Glu Glu Asp Glu Leu Glu Ser Gly Asp Gln
                935                 940                 945
Glu Asp Glu Asp Asp Glu Ser Glu Asp Pro Gly Lys Asp Leu Asp
                950                 955                 960
His Ile
```

<210> 9
<211> 368
<212> PRT
<213> Homo sapiens

<220> -
<223> 2914719

<400> 9

```
Met Ser Leu Phe Gly Thr Thr Ser Gly Phe Gly Thr Ser Gly Thr
  1               5                  10                  15
Ser Met Phe Gly Ser Ala Thr Thr Asp Asn His Asn Pro Met Lys
                 20                  25                  30
Asp Ile Glu Val Thr Ser Ser Pro Asp Asp Ser Ile Gly Cys Leu
                 35                  40                  45
Ser Phe Ser Pro Pro Thr Leu Pro Gly Asn Phe Leu Ile Ala Gly
                 50                  55                  60
Ser Trp Ala Asn Asp Val Arg Cys Trp Glu Val Gln Asp Ser Gly
                 65                  70                  75
Gln Thr Ile Pro Lys Ala Gln Gln Met His Thr Gly Pro Val Leu
                 80                  85                  90
Asp Val Cys Trp Ser Asp Asp Gly Ser Lys Val Phe Thr Ala Ser
                 95                 100                 105
Cys Asp Lys Thr Ala Lys Met Trp Asp Leu Ser Ser Asn Gln Ala
                110                 115                 120
Ile Gln Ile Ala Gln His Asp Ala Pro Val Lys Thr Ile His Trp
                125                 130                 135
Ile Lys Ala Pro Asn Tyr Ser Cys Val Met Thr Gly Ser Trp Asp
                140                 145                 150
Lys Thr Leu Lys Phe Trp Asp Thr Arg Ser Ser Asn Pro Met Met
                155                 160                 165
Val Leu Gln Leu Pro Glu Arg Cys Tyr Cys Ala Asp Val Ile Tyr
                170                 175                 180
Pro Met Ala Val Val Ala Thr Ala Glu Arg Gly Leu Ile Val Tyr
                185                 190                 195
Gln Leu Glu Asn Gln Pro Ser Glu Phe Arg Arg Ile Glu Ser Pro
                200                 205                 210
Leu Lys His Gln His Arg Cys Val Ala Ile Phe Lys Asp Lys Gln
                215                 220                 225
Asn Lys Pro Thr Gly Phe Ala Leu Gly Ser Ile Glu Gly Arg Val
                230                 235                 240
Ala Ile His Tyr Ile Asn Pro Pro Asn Pro Ala Lys Asp Asn Phe
                245                 250                 255
Thr Phe Lys Cys His Arg Ser Asn Gly Thr Asn Thr Ser Ala Pro
                260                 265                 270
Gln Asp Ile Tyr Ala Val Asn Gly Ile Ala Phe His Pro Val His
                275                 280                 285
Gly Thr Leu Ala Thr Val Gly Ser Asp Gly Arg Phe Ser Phe Trp
```

```
                    290                      295                      300
Asp Lys Asp Ala Arg Thr Lys Leu Lys Thr Ser Glu Gln Leu Asp
                    305                      310                      315
Gln Pro Ile Ser Ala Cys Cys Phe Asn His Asn Gly Asn Ile Phe
                    320                      325                      330
Ala Tyr Ala Ser Ser Tyr Asp Trp Ser Lys Gly His Glu Phe Tyr
                    335                      340                      345
Asn Pro Gln Lys Lys Asn Tyr Ile Phe Leu Arg Asn Ala Ala Glu
                    350                      355                      360
Glu Leu Lys Pro Arg Asn Lys Lys
                    365
```

<210> 10
<211> 1527
<212> DNA
<213> Homo sapiens

<220> -
<223> 144861

<400> 10

```
ctcgccccga gagagacccg gccatgcagg agccgctgct gggagccgag ggcccggact   60
acgacacctt ccccgagaag ccgcccccgt cgccagggga cagggcgcgg gtcgggaccc  120
tgcagaacaa aagggtgttc ctggccacct tcgccgcagt gctcggcaat ttcagctttg  180
ggtatgccct ggtctacaca tcccctgtca tcccagccct ggagcgctcc ttggatcctg  240
acctgcatct gaccaaatcc caggcatcct ggtttgggtc cgtgttcacc ctgggagcag  300
cggccggagg ctgagtgcca tgatcctcaa cgacctcctg ggccggaagc tgagcatcat  360
gttctcagct gtgccgtcgg cggccggcta tgcgctcatg gcgggtgcgc acggcctctg  420
gatgctgctg ctcggaagga cgctgacggg cttcgccggg ggtctcacag ctgcctgcat  480
cccggtgtac gtgtctgaga ttgctccccc aggcgttcgt ggggctctgg gggccacacc  540
ccagctcatg gcagtgttcg gatccctgtc cctctacgcc cttggcctcc tgctgccgtg  600
gcgctggctg gctgtggccg gggaggcgcc tgtgctcatc atgatcctgc tgctcagctt  660
catgcccaac tcgccgcgct tcctgctctc tcggggcagg gacgaagagg ccctgcgggc  720
gctggcctgg ctgcgtggga cggacgtcga tgtccactgg gagttcgagc agatccagga  780
caacgtccgg agacagagca gccgagtatc gtgggctgag gcacgggccc cccacgtgtg  840
ccggcccatc accgtggcct tgctgatgcg cctcctgcag cagctgacgg gcatcacgcc  900
catcctggtc tacctgcagt ccatcttcga cagcaccgct gtcctgctgc cccccaagga  960
cgacgcagcc atcgttgggg ccgtgcggct cctgtccgtg ctgatcgccg ccctcaccat 1020
ggacctcgca ggccgcaagg tgctgctctt cgtctcaggc cacctgtctg tccacagcgg 1080
ccatcatgtt tgctgccaac ctgactctgg ggctgtacat ccactttggc cccaggcctc 1140
tgagccccaa cagcactgcg ggcctggaaa gcgagtcctg gggggacttg gcgcagccct 1200
tgcagcaccc gctggctatc tcaccctggt gcccctgctg gccaccatgc tcttcatcat 1260
gggtaggtgt ggtggtggct cagaggggca ggctgtcttg ggtgttaagg gatgggtgat 1320
gtgtctgggg tggctggaga ggggggtctc cagcagctca gcggagacag acacagccgc 1380
ctccagtcac cccacagggc ctgaacctgc ctcctcctcc gagcaggcta cgccgtgggc 1440
tggggtccat cacctggctg ctcatgtctg aggtcctgcc ctgcgtgccg tggcgtggct 1500
caaggcctac gccgtgggct ggggtcc                                    1527
```

<210> 11
<211> 1270
<212> DNA
<213> Homo sapiens

<220> -

<223> 607812

<400> 11
ggatgccatc cttctcaaaa gacttattga cagtgccaaa gctcggtact ggacacaacg 60
agggacctgg gtctacgata acgcgctttt gctcctcctg aagtgtcttt ggtccaacgt 120
tgttccagag tgtaccatgg cttccagtaa cactgtgttg atgcggttgg tagcctccgc 180
atattctatt gctcaaaagg caggaatgat agtcagacgt gttattgctg aaggagacct 240
gggtattgtg gagaagacct gtgcaacaga cctgcagacc aaagctgacc gattggcaca 300
gatgagcata tgttcttcat tggccccggaa attcccccaaa ctcacaatta taggggaaga 360
gctcgtggtc tgggttgatc ctctggatgg aaccaaggaa tataccgaag gtcttcttga 420
caatgtaaca gttcttattg gaattgctta tgaaggaaaa gccatagcag gagttattaa 480
ccagccatat tacaactatg aggcaggacc agatgctgtg ttggggagga caatctgggg 540
agttttaggt ttaggcgcct ttgggtttca gctgaaagaa gtccctgctg ggaaacacat 600
tatcacaact actcgatccc atagcaacaa gttggttact gactgtgttg ctgctatgaa 660
ccccgatgct gtgctgcgag taggaggagc aggaaataag attattcagc tgattgaagg 720
caaagcctct gcttatgtat ttgcaagtcc tggttgtaag aagtgggata cttgtgctcc 780
agaagttatt ttacatgctg tgggaggcaa gttaaccgat atccatggga atgttcttca 840
gtaccacaag gatgtgaagc atatgaactc tgcaggagtc ctggccacac tgaggaatta 900
tgactactat gcaagccgag ttccagaatc tattaaaaat gcacttgttc cttaaaggaa 960
agtttcattt ggccgggcgc ggtggctcat gcctgtagtc ccggcacttt gggaggccga 1020
ggcaggtgga tcacttgagc tcgggagttt gagaccagcc tgggcaatat cgtgagaccc 1080
catctctaca aaaatacaaa ttaactgggc atcctgtcat gcgcctgtcg tcccagctac 1140
ttgagaggct gaagcagaag aatctcttgg gcccgggagg cggaggttgc agtgagctga 1200
gatcgtgcca ctgcactcca gcctgagtga caggagttaa gccctgtctc agaaatagaa 1260
tataaaaggg 1270

<210> 12
<211> 902
<212> DNA
<213> Homo sapiens

<220> -
<223> 1259384

<400> 12
gcggaatgca gcggcccgag gcctggccac gtccgcaccc gggggagggg gccgcggcgg 60
cccaggccgg gggcccggcg ccgcctgctc gagccgggga gccctcgggg ctgcggttgc 120
aggaaccttc cctctacacc atcaaggctg ttttcatcct agataatgac gggcgccggc 180
tgctggccaa gtattatgat gacacattcc cctccatgaa ggagcagatg gttttcgaga 240
aaaatgtctt caacaagacc agccggactg agagtgagat tgcatttttt gggggtatga 300
ccatcgtcta caagaacagc attgacctct tcctatacgt ggtgggctca tcctacgaga 360
atgagctgat gctcatgtct gttctcacct gcctgtttga gtctctgaac cacatgttaa 420
ggaagaacgt ggagaagcgc tggttgctgg agaacatgga cggagccttc ttggtgctgg 480
acgagattgt ggatggcggt gtgattctgg agagtgaccc ccagcaagtg atccagaagg 540
tgaattttag ggcagatgat ggcggcttga ctgaacagag tgtggcccag gttcttcagt 600
ctgccaagga acaaattaaa tggtcgttat tgaaatgaag gctgtggatt caaggctccc 660
tgccccccag atcatttccc caatcctggc aaaagcccaa agatcccagg gtcaggagag 720
acccctctgt atccccaggt ccctccccaga actgactcct aaggtctcca gccagggctt 780
ctgagatgca aaggtttggc ctcaggagag tcacctttc tcacggccct ggccttaact 840
catatcttag gcattcctgg ccccagggcc ctaataaacc tgcttttgtc ttctgcaaat 900
aa 902

<210> 13
<211> 2026
<212> DNA
<213> Homo sapiens

<220> -
<223> 1340813

<400> 13
gcagaggcgg cgagaggcgg gctgaggcgg cccagcggcg gcaggtgagg cggaaccaac 60
cctcctggcc atgggagggg ccgtggtgga cgagggcccc acaggcgtca aggcccctga 120
cggcggctgg ggctgggccg tgctcttcgg ctgtttcgtc atcactggct tctcctacgc 180
cttccccaag gccgtcagtg tcttcttcaa ggagctcata caggagtttg ggatcggcta 240
cagcgacaca gcctggatct cctccatcct gctggccatg ctctacggga caggtccgct 300
ctgcagtgtg tgcgtgaacc gctttggctg ccggcccgtc atgcttgtgg ggggtctctt 360
tgcgtcgctg ggcatggtgg ctgcgtcctt ttgccggagc atcatccagg tctacctcac 420
cactggggtc atcacggggt tgggtttggc actcaacttc cagccctcgc tcatcatgct 480
gaaccgctac ttcagcaagc ggcgccccat ggccaacggg ctggcggcag caggtagccc 540
tgtcttcctg tgtgccctga gcccgctggg gcagctgctg caggaccgct acggctggcg 600
gggcggcttc ctcatcctgg gcggcctgct gctcaactgc tgcgtgtgtg ccgcactcat 660
gaggcccctg gtggtcacgg cccagccggg ctcggggccg ccgcgaccct cccggcgcct 720
gctagacctg agcgtcttcc gggaccgcgg ctttgtgctt tacgccgtgg ccgcctcggt 780
catggtgctg gggctcttcg tcccgcccgt gttcgtggtg agctacgcca aggacctggg 840
cgtgcccgac accaaggccg ccttcctgct caccatcctg ggcttcattg acatcttcgc 900
gcggccggcc gcgggcttcg tggcggggct tgggaaggtg cggccctact ccgtctacct 960
cttcagcttc tccatgttct tcaacggcct cgcggacctg gcgggctcta cggcgggcga 1020
ctacggcggc ctcgtggtct tctgcatctt ctttggcatc tcctacggca tggtgggggc 1080
cctgcagttc gaggtgctca tggccatcgt gggcaccccac aagttctcca gtgccattgg 1140
cctggtgctg ctgatggagg cggtggccgt gctcgtcggg ccccccttcgg gaggcaaact 1200
cctggatgcg acccacgtct acatgtacgt gttcatcctg gcgggggccg aggtgctcac 1260
ctcctccctg attttgctgc tgggcaactt cttctgcatt aggaagaagc ccaaagagcc 1320
acagcctgag gtggcggccg tggaggagga gaagctccac aagcctcctg cagactcggg 1380
ggtggacttg cgggaggtgg agcatttcct gaaggctgag cctgagaaaa acgggggaggt 1440
ggttcacacc ccggaaacaa gtgtctgagt ggctgggcgg ggccggcagg cacagggagg 1500
aggtacagaa gccggcaacg cttgctattt attttacaaa ctggactggc tcaggcaggg 1560
ccacggctgg gctccagctg ccggcccagc ggatcgtcgc ccgatcagtg ttttgagggg 1620
gaaggtggcg gggtgggaac cgtgtcattc cagagtggat ctgcggtgaa gccaagccgc 1680
aaggttacaa ggcatcctca ccaggggccc cgcctgctgc tcccaggtgg cctgcggcca 1740
ctgctatgct caaggacctg gaaacccatg cttcgagaca acgtgacttt aatgggaggg 1800
tgggtgggcc gcagacaggc tggcagggca ggtgctgcgt ggggccctct ccagcccgtc 1860
ctaccctggg ctcacatggg gcctgtgccc acccctcttg agtgtcttgg ggacagctct 1920
ttccacccct ggaagatgga aataaacctg cgtgtgggtg gagtgttagg aaaaaaaaaa 1980
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aagtct 2026

<210> 14
<211> 2829
<212> DNA
<213> Homo sapiens

<220> -
<223> 1689731

<400> 14
ccgatcccaa gttccagcta gagaaggtgg gcggcagcgc cagcccggcg ccccctccca 60
gctcccagcg atttgagagg gaagatctga tctcctagaa ttagagttgg tacagaaacc 120
atttcagctc caaaaatgtt ggaggaagat atggaagtcg ccataaagat ggtggttgta 180
gggaatggag cagttggaaa atcaagtatg attcagcgat attgcaaagg catttttaca 240
aaagactaca agaaaaccat tggagttgat tttttggagc gacaaattca agttaatgat 300
gaagatgtca gactaatgtt atgggacact gcaggtcagg aggaatttga tgcaattaca 360
aaggcctact atcgaggagc ccaggcttgt gtgctcgtgt tctctaccac agatagggaa 420
tcttttgaag cagtttccag ttggagagag aaagtagtag ccgaagtggg agatatacca 480
actgtacttg tgcaaaacaa gattgatctt ctggatgatt cttgtataaa gaatgaggaa 540

gctgaggcac tggcaaaaag gttaaagtta agattctaca gaacatcagt gaaagaagat 600
ctaaatgtga atgaagtttt taagtatttg gctgaaaaat accttcagaa actcaaacaa 660
caaatagctg aggatccaga actaacgcat tcaagtagta acaagattgg tgtctttaat 720
acatctggtg gaagtcactc cggtcagaat tcaggtaccc tcaatggtgg agatgtcatc 780
aatcttagac ccaacaaaca aaggaccaag aaaaacagaa atccttttag cagctgtagc 840
ataccctaag atgttttggg aggaaaacaa ttgaattaca ttgtgcaatg cattaagaaa 900
cccatccagc tgtcatggta tgttacaagg ttttcagcag actttgctcc taatggctct 960
ctgaaagtga cagacttaaa tattgctctg cagtgctttt cagaatgtag agtgagacct 1020
ctggtggaaa aattattgcc ctgtggactc ctttaatttt ttttacatta gacgaagctt 1080
ctcctgtttg taaaggaatg ctataagaaa tttcaaacac aggttttgcg gaattttaaa 1140
tgctggaaaa cataaatgaa atgcctaaat atattgttac agattttaat actgtggatc 1200
aaggaaagat agcaagcatt cttttcctga aattggtcat gtagcttttc tgcgaaatag 1260
tattgaatct gaatactttc acaaaataat actgaaaggt aaaccctaaa tcttgccaga 1320
ctgcattagt attttggcat tgaaacttgt gcaatatctg tgtaatgtag aatgtggggt 1380
gtgcatgcag atgtctggta ctggagtaag ctgaaggtat gtgcagtatt taagaaaatg 1440
tacattctaa ttaattcact tataaaacca tttacatatt gtgtatatgt ttttattttc 1500
aaatgataaa ctctagacat atattttgca aaatgagctt gggtttaaaa agtgctttgt 1560
gccaaaaata tcatacctaa tttttatagt gctttactat cagaatttat attgtttatt 1620
atatctttta aaagcagtaa gcagttttca tctgacagcg tagataaaag aggagtatag 1680
gaatttgaat gctagagcat ctcccaggtg agccttattg cagatctttt tgatatagtc 1740
acccataagc atggtgagct cctgctctca taaggtcagc caaagcccct taagaaggta 1800
cacagcacat ctgcagtcat tcaggaggtc ctgtgaagtg tggtgatgca tctaacacct 1860
tttgctgctc actgagaacc cagctttgaa cttacaagat tacactaacc atttttaaag 1920
tgcacaaaaa tattattatt agtcttaact aaaattacat attttataat gccctgcaac 1980
tctaaaccat cctttttgttt acacattttt gtatggtatt ttaagattat caatcacttc 2040
aacatttcta taaccttgtt tgtatagcac tttctcaaga gctgtcagtc ccacatacag 2100
aatcatttct tttacaaaat gatttaatat gtaatcatgt gtcaggacat aaattttgtc 2160
acttgggaca gtagctttcc agctaagata attgcgtgaa ctctgttaaa actgattatc 2220
actcatcaat ggctttattt ttaacccata aataacccct ttcaatatac ctggttttta 2280
cagcattgca aaatctgtaa aaataacttt tattggaaaa tgtcatgaag tcatttttca 2340
ttatttgtac taaacaatgt tttatgtact ccttttgtgt tttatctttt gtttttccat 2400
ctgtactccc ctcatcctcc taatgttgag aaaatcaaat ttcctgggtg tgtccttgtt 2460
aagtgatcaa aagtaaatgt tctctttgcc attttagatg gaaaaataaa aaaaattaaa 2520
aagtaaattt ttgcttttca ttatattgaa atgatctggc attatctatt tcacagtaat 2580
gtaaaatgct tatttttatg tcaccagttg tcaccactcc cattgctgat agtgacatag 2640
gttttggggg caaaattcag ttcctaacaa acagtggata ctgctaatga gtcatctttt 2700
ctactctctt caattttcta ctctctactc tcctctttaa tttttgcaca aaatcctact 2760
ttgcaacaaa tagtcttttg tccttattgg gaggaaaatc acttgctaaa tgtgctctag 2820
ttagttctt 2829

<210> 15
<211> 1589
<212> DNA
<213> Homo sapiens

<220> -
<223> 2751730

<400> 15
cccggctgcc agcaccatgt ccgccagggg agattttggg aatccactga gaaaattcaa 60
gttggtgttc ttgggggagc agagcgtcgg gaagacgtct ctgattacga ggttcatgta 120
cgacagcttc gacaacacat accaggcaac cattgggatt gacttcttgt caaaaaccat 180
gtacttggag gaccgcacgg tgcgactgca gctctgggac acagctggtc aggagaggtt 240
ccgcagcctg atccccagct acatccggga ctccacggtg gctgtggtgg tgtacgacat 300
cacaaatctc aactccttcc aacagacctc taagtggatc gacgacgtca ggacagagag 360
gggcagtgat gttatcatca tgctggtggg caacaagacg gacctggctg ataagaggca 420
gataaccatc gaggaggggg agcagcgcgc caaagaactg agcgtcatgt tcattgagac 480

cagtgcgaag actggctaca acgtgaagca gctttttcga cgtgtggcgt cggctctacc 540
cggaatggag aatgtccagg agaaaagcaa agaagggatg attgacatca agctggacaa 600
accccaggag cccccggcca gcgagggcgg ctgctcctgc taatgcagag ccgacctgtg 660
gcttcccatg acactccttg cttgttgtgt tgcttcctat tggctagctt cctaaggggg 720
gagggaaccg agttatcaag atgggaggat tttcttttc tctctgtctt taggagtagg 780
gtgggatggg gagggaggct gggcatcagg gatcacatca ctcttaacgg ctgttactta 840
aacaactatt ttttggtttg gttgtaatat attgtacttt attaagattg ccaaaaactg 900
ttaaaattta aaaaaaattt aaatcatgtg tatacaattt tttgccagat aaaaatgtag 960
tcatttttat ttgaaagatg tgctttttgt ttttgtatat ttgtaaactt atagagaacc 1020
ttttccacac acctcctcct tcctgttctc tttgaaccat tcatcacctc tgccttcctc 1080
ctatccccag cccaataaat taaaacaatt aactgagcaa attaattagg cttcagtctg 1140
gggccatctg gccccactct ctaggcccta ctccagttaa atcaaacatt ggttgaacac 1200
atcagcctct gaaaaggtag ctctgactct tgtctttcca tggccagagt gggtcgtcaa 1260
cctcgtgtgg agctgagacc aaggtctgtg ggctctaggc agcccggcag cagcccccag 1320
aggaactgct cccacctctc cttctctttc cctgaaaagc cttgaagggg gttgagggta 1380
caggcaacct tgtgttctct gtgccctcag ggactggcag agcaagaggc catcagaagg 1440
acgaggtgct gttggcaatt gcagctgcgt ggtgcgtggt ttatagactg ggtgccatg 1500
tgtacttctg aaaatatttt aaccagccca cccagcagtc agtacacatg cctgggccaa 1560
gtctaagaaa ctcagtgtgt gatgttggg 1589

<210> 16
<211> 2533
<212> DNA
<213> Homo sapiens

<220> -
<223> 2794975

<400> 16
cacagcgact gcgcggacgg gttcctgagt ggaacatggc gacttgcgcc gaaatcctgc 60
ggagcgagtt ccccgaaatt gacggacaag tcttcgacta cgtgaccggc gtcttgcaca 120
gcggcagcgc ggacttcgag tctgtggatg acctggtgga agctgtaggg gaactattgc 180
aagaggtgtc cggggacagc aaggatgacg cgggcatcag ggccgtgtgc cagcgcatgt 240
acaacactct gcgtctggct gagccacaaa gccagggaaa tagccaggtg ctactggacg 300
cccctatcca gttgtcaaag ataacggaga actacgactg tggaaccaaa cttccaggac 360
tgctaaagag ggaacagtcc tcgacagtga atgcaaagaa gttagagaag gccgaggctc 420
gacttaaggc aaagcaggag aagcgctcag agaaggacac gctcaagacc agcaaccctc 480
tagtcttaga agaggcatca gccagccagg caggcagcag aaaggagagt cggttggaat 540
catctggcaa gaacaaatcc tatgatgtgc gaattgagaa ctttgatgtg tcttttggcg 600
atagagtact gctggctgga gcggatgtga acctggcatg ggccgccgt tacgggctgg 660
tggggcggaa tgggttgggg aagacaacgt tactgaagat gctggccacc cggagtctgc 720
gggttccagc ccacatttcc ctgctgcacg ttgagcaaga ggttgctgga gatgacactc 780
ctgccctgca gagtgtgctg gagagtgaca gtgtgcgaga ggatttgcta cggagggagc 840
gggagctcac tgcccagatt gctgctggca gggcggaggg ctctgaagct gcagagctgg 900
cagaaatcta tgccaaactg gaggagattg aggctgacaa ggcacctgcc agggcatcag 960
tcattctcgc tgggcttggc tttaccccta aaatgcagca gcagcccacc cgggagttct 1020
caggtggctg gaggatgagg ctggccctgg cccgggccct ctttgctagg ccagatcttc 1080
tgctgttaga tgaacctaca aacatgctgg atgtcagggc catcctgtgg ctggagaatt 1140
acctgcagac gtggccctcc accatcctag tcgtctccca cgaccgcaac ttcttgaatg 1200
ccatcgccac agacatcatc cacctgcaca gccagcggct agatggttac cggggagact 1260
ttgagacctt catcaagagt aagcaggagc ggctgctcaa ccagcagcgt gaatatgagg 1320
cgcagcagca gtatcgccag cacatccagg ttttcattga ccggtttcgc tacaatgcca 1380
acagggcctc tcaagtgcag agtaaactca agatgctgga gaagctgcct gagctgaagc 1440
ctgtggacaa ggaatcagag gtcgtaatga agttccctga tgggtttgag aagttctcgc 1500
cgccaattct gcagctagat gaggtggatt tctactacga tccgaagcac gtcatcttca 1560
gtcgcctctc tgtgtctgct gatctcgagt ctcgcatctg tgtggttgga gagaatgggg 1620
ctgggaagtc taccatgctg aagctgcttt tgggggacct ggcacctgtt cggggcatca 1680

gacacgctca caggaatctg aagattggct atttcagcca gcaccatgtg gagcagctgg 1740
acctaaacgt cagtgctgtg gaactgctgg cacgcaagtt tcctgggcgg cctgaggagg 1800
agtaccgtca ccagctgggt cggtatggca tctccggaga actggccatg cgtcctcttg 1860
ccagcctgtc tgggggccag aagagccgag tggcctttgc tcagatgact atgccctgcc 1920
ccaacttcta cattctggat gaacccacaa accacctgga catggagacc attgaggctc 1980
tgggccgtgc cctcaacaat ttcaggggtg gtgtgattct ggtgtcccac gatgagcgct 2040
ttatcaggct ggtgtgccgg gagttgtggg tatgcgaagg aggcggcgtc acccgtgtgg 2100
aaggaggatt tgaccagtac cgcgccctcc tccaggaaca gttccgccgc gaaggcttcc 2160
tctagggcca ccaggctgag gactcgccca ggacatggac tggtctctca gacccctggg 2220
ccaccatgta ggccaccact ccaggccgtg gacttccccc aacttgggga cagccttatt 2280
cccaaatgtc tctatccttt tgactggagc atcttctgca caaccttggg agcccatcca 2340
agggttggtg aggactggtc tcccgggggt gggggtctgg ggggtaccct ctggggttat 2400
agattccccc actgccccag ctctgactgg accccaagtg gctgctatgt aaattaaatc 2460
tctccccgcg tcaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2520
aaaaaaaaaa aag 2533

<210> 17
<211> 4312
<212> DNA
<213> Homo sapiens

<220> -
<223> 2797710

<400> 17
cttcgagccg ccacgtaatg ccacgtcccc gcgcatgcgc atcttggccg ctgctggcgg 60
ctgtttccgg gcttagaggg ctggagtggc cgccgagttg gaggcggtgg tggcagcagt 120
aggagtgtgt agagtgcggg attgggggcc aggccctgcg gagggcgggg gaagttgtct 180
tctttttttt ccggaggggc cggtaaacct ggtggctgaa cggcaagatg aatttcctcc 240
gcggggtaat ggggggtcag agtgccggac cccagcacac agaagccgag acgattcaaa 300
agctttgtga cagagtagct tcatctactt tattggatga tcgaagaaat gctgttcgtg 360
ctctcaaatc attatctaag aaataccgct tggaagtggg tatacaagct atggaacatc 420
ttattcatgt tttacaaaca gatcgttcag attctgaaat aataggttat gctttggaca 480
cactatataa tataatatct aatgaagaag aagaagaagt agaagaaaat tccacaagac 540
agagtgaaga tttgggaagc caatttacag aaattttcat taagcagcag gaaaatgtca 600
ctcttctgtt atctttattg gaggagtttg atttccatgt ccgctggcct ggtgtgaagc 660
ttcttacttc tcttttaaaa caactagggc ctcaggtgca acaaattatt ttagtcagtc 720
ctatgggtgt ttcaagattg atggacttac tagcggattc cagggaagtt atacgtaatg 780
atggcgtctt actactgcag gcactaacaa gaagcaatgg tgcaatccag aaaattgttg 840
cttttgaaaa tgctttcgag agactactgg acattatttc agaggagggg aacagtgatg 900
gaggtatagt agttgaagat tgtttgattt tgctccaaaa cttattaaaa aacaacaact 960
ccaatcgaaa ttttttaaa gaaggctcat atattcaacg tatgaaacct tggtttgaag 1020
ttggagatga aaattctggc tggtctgcac agaaagtgac caatctacat ctaatgctac 1080
agcttgttcg tgtattggta tctcccacca accctcctgg tgctaccagt agctgccaga 1140
aggctatgtt ccagtgtggg ttattgcagc agctttgtac tatcctaatg gctactgggg 1200
ttcctgctga tatcctgact gagaccataa atactgtatc agaagttatt cgaggctgcc 1260
aagtaaacca agactacttt gcatctgtaa atgcaccttc aaacccacca agaccggcaa 1320
ttgtagtact tctcatgtcc atggttaatg aaaggcagcc atttgttttg cgctgtgctg 1380
ttctctattg tttccagtgt ttcttgtata aaaaccaaaa aggacaagga gaaatcgtgt 1440
caacactttt accttctacc attgatgcaa caggtaattc agtttcagct ggccagttat 1500
tatgtggagg tttgttttct actgattcac tttcaaactg gtgtgctgct gtggcccttg 1560
cccatgcgtt gcaagaaaat gccacccaga aagaacagtt gctcagggtt caacttgcta 1620
caagtattgg caaccctcca gtttctttac ttcaacagtg caccaatatt ctttcacagg 1680
gaagcaaaat acaaacaaga gttggattat taatgttgct ttgtacctgg ctaagcaatt 1740
gtcccattgc agtaacgcat tttcttcaca attcagccaa tgttccattt cttacaggac 1800
aaattgcaga aaatcttgga gaagaagagc agttggtcca aggcttatgt gcccttttgt 1860
tgggcatttc gatttatttc aatgataact cacttgagag ctacatgaaa gagaagctaa 1920

```
aacaactgat tgagaagagg attggcaaag agaatttcat agagaaacta ggatttatta 1980
gcaaacatga gttgtattcc agagcatctc agaaacccca gccaaacttt cccagtccag 2040
aatacatgat atttgatcat gagtttacga agctggtaaa agaacttgaa ggtgttataa 2100
ctaaggctat ttataagtcc agtgaagaag ataaaaaaga agaagaggtg aaaaaaacat 2160
tagaacagca tgacaatatt gtgactcact acaaaaatat gattcgagag caggatctcc 2220
aacttgagga attaaggcag caggtttcta cattaaaatg tcaaaatgaa cagctccaga 2280
cggcagtcac acagcaagta tcacagatcc agcagcacaa agaccagtat aatcttctta 2340
aaatacagct aggaaaagac aatcagcatc aaggttctta cagtgagggg gctcagatga 2400
atggcattca gccagaagaa attggtagat tgcgagaaga gatagaagaa ttaaaacgta 2460
atcaggaact tttacaaagc cagctgactg aaaaggactc tatgattgaa aatatgaaat 2520
cttcccaaac atctggcaca aatgaacagt cttcagcaat agtttcagct agagattctg 2580
aacaagttgc agaattaaaa caggaactgg caactttaaa gtctcagtta aactcacaat 2640
ctgtggagat caccaaacta cagacagaaa agcaggaact gttacagaaa acagaagcgt 2700
ttgcaaaatc agttgaggta caaggagaga ccgagactat aatagccacc aaaactactg 2760
atgtagaagg aagactgtca gcattattac aagagaccaa agagttaaag aatgaaatta 2820
aagctctgtc tgaggaaaga actgccatta aagagcagct ggattcatct aatagtacca 2880
ttgccatttt acaaactgag aaagacaaac tagagttgga aattacagat tctaaaaaag 2940
aacaagatga tctcttggtg ctcttggccg atcaagatca gaaaatactg tcattgaaga 3000
ataaactcaa ggatcttggt catccagttg aagaagagga tgaacttgaa tctggagacc 3060
aagaggatga ggatgatgaa agtgaagatc ctggcaagga tctagatcat atctagtttt 3120
caaatctcta ggaacaatag agattatatc ataactctga agatggactc taaactttgg 3180
tttacctcca tggaaaataa agatttagaa accaggtgga aacattcac tattaagact 3240
attgatatat ttttgtaatg ttgccaccca tgttgaaaac cttaaaactg aatttatgta 3300
gaacacacat cttttattta aatccatctg aactgtccca aaatgtaatt tgcaaagagc 3360
ttcaaacacc aaaaatatac cattttaagg catgtggtga caatgctatt tgccttagtt 3420
cattcgttgt ttactttgca aaatttgtga cttttaagta ctaacacatt tgattgaatg 3480
tgtattgaat tggaattatt atttctaaat tcagcatttg agtccttagt tgatgatttt 3540
cttttaaaga aattagttta tttaaaagca agatagcctg ggttgggagt tagtgcatat 3600
catcagccct ccatctaatg tgattgctaa gatgtgcagg ataggacagc ttagaatgag 3660
tgccttgaca cctacatatg gctgcattcc atacctcagt gaagaagccg aaggatgctg 3720
acacacagtg atgtggaggg cattccggag aaacttttgc aacagtgtat taatgtgatt 3780
ggtttctttt ttgttttaag ctaaacatct aaaacaaatt tcaagttaac agttcataca 3840
actcagattt agtattattg tttatcataa aacatattat gttcctagtt ttaaaaagac 3900
ccaaagtatg attaatataa tttattggca tttttgctga ataggtttaa gtcagataat 3960
agattttaaa aaagcaaatg aggcaatgtg tcaaaattta atgtttcata ataaaaatag 4020
atacgttgac ttaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4080
aaaaaaaaaa aaaaaaaaaa aagggggggc cgctctaggg gtcccaggtt taggtagggg 4140
tgaatgggag gtcaaggccc ttctaagggg tcccctaatt tcatttcacg ggcggcggtt 4200
ttaaaaggcc gtgactggga aacccctggg gttacccaat ttaatcgcct tggggaaaac 4260
cccctttgg caagcgggcg taatagcgaa ggggcccgca cggttcgcct tt         4312
```

```
<210> 18
<211> 2370
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 2354
<223> a or g or c or t, unknown, or other

<220> -
<223> 2914719

<400> 18
gtttctaccg caggcttaag gaggcttcgg gctcctggga tttctgtccg cgctcctggc  60
cctcgtcctt cgcgccagag caggttcgca aactcctcag acccttctgc tcccggccgc 120
```

cgctttccgc cggggcgaga cccccaggtt caaaatgagc ctgtttggaa caacctcagg 180
ttttggaacc agtgggacca gcatgtttgg cagtgcaact acagacaatc acaatcccat 240
gaaggatatt gaagtaacat catctcctga tgatagcatt ggttgtctgt cttttagccc 300
accaaccttg ccggggaact ttcttattgc aggatcatgg gctaatgatg ttcgctgctg 360
ggaagttcaa gacagtggac agaccattcc aaaagcccag cagatgcaca ctgggcctgt 420
gcttgatgtc tgctggagtg acgatgggag caaagtgttt acggcatcgt gtgataaaac 480
tgccaaaatg tgggacctca gcagtaacca agcgatacag atcgcacagc atgatgctcc 540
tgttaaaacc atccattgga tcaaagctcc aaactacagc tgtgtgatga ctgggagctg 600
ggataagact ttaaagtttt gggatactcg atcgtcaaat cctatgatgg ttttgcaact 660
ccctgaaagg tgttactgtg ctgacgtgat ataccccatg gctgtggtgg caactgcaga 720
gaggggcctg attgtctatc agctagagaa tcaaccttct gaattcagga ggatagaatc 780
tccactgaaa catcagcatc ggtgtgtggc tatttttaaa gacaaacaga acaagcctac 840
tggttttgcc ctgggaagta tcgaggggag agttgctatt cactatatca acccccccgaa 900
ccccgccaaa gataacttca cctttaaatg tcatcgatct aatggaacca acacttcagc 960
tcctcaggac atttatgcgg taaatggaat cgcgttccat cctgttcatg gcacccttgc 1020
aactgtggga tctgatggta gattcagctt ctgggacaaa gatgccagaa caaaactaaa 1080
aacttcggaa cagttagatc agcccatctc agcttgctgt ttcaatcaca atggaaacat 1140
atttgcatac gcttccagct acgactggtc aaagggacat gaattttata atccccagaa 1200
aaaaaattac atttcctgc gtaatgcagc cgaagagcta aagcccagga ataagaagta 1260
gtggctggag actctggctc agccagagtt gtttctctcc actctgcctc atctctgtac 1320
gaatttgggt cccagccttg ttgggttgtc agccatggac atggatttca acccctggag 1380
aaaacgatgt cattgttcag cagctgagag cccaggcgtc cgcggcgact tgccgtctct 1440
ccattccact gcctgttgca gagtttttct gtaactaagg gggttgaggt tattgtagac 1500
gttagattgc gggcaccgcc agggattttg cagcgcttca gtgtacgtgt tagagaatat 1560
tggaaaagcg tctgtgagcc ccgtgctgta ttttgtaata aagtcttttg cagattgctt 1620
cccgagcttc ctttgtcctt ttctcccctt gcccaccccg taacctcagg aacatgcgtc 1680
ctgcccagca tcagcgtggg gttttgagtt gagatttcag acaccctctg ggaaatgcgg 1740
caacccttagg ggaaagggag tccccagctg cgctcacttc tgctgcgtgg aacggcagcc 1800
tctgtgagcc ctggtgggca gagtttgaat gtgtttttcc ttgcttccct cattcccatc 1860
ttcaaaatcc ccagtgcttt ctggccttgc tgctcagatt tccgagtgac tcaaatgggg 1920
actgttactt gtgctgggtg acaggccatt tgtgggtaac ctcctaaggc ccaagtgggt 1980
gacacttgcg tgactttcaa gttagaaccc aagccccctg catgggaatt gccctgaact 2040
cttacccacc ccggtcccct gcatgggaat tgccctgaac cctcagtagt ggttgtttgg 2100
ctgttctttt gtattttgtg taatttaaaa ttcctggttg gtaagtattc aaaatgaaat 2160
tcagctgggc tgagaaaaag tgtgactttt gggtctgtca ctgtatttct cacctgtgat 2220
ctcaaatgct tcttaggcct ttctgtttgg actaatgtgt gaagtctgac ttgctgagtg 2280
taaaattcta gtatcgatag tgttgtaaga tgtgtttgcc tactcataaa aacaatgaa 2340
aataaaattt catnctgcaa tgttagttgg 2370